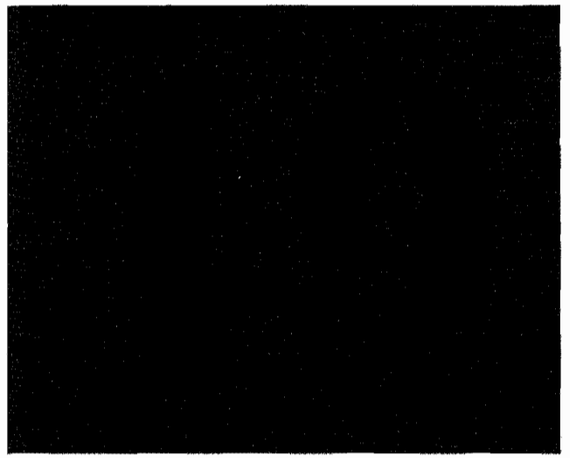


Razonamiento con incertidumbre

A menudo, un agente sólo tiene información incierta acerca de su tarea y del entorno. Las técnicas que acabamos de describir hasta ahora, tenían ciertas limitaciones a la hora de representar y razonar con conocimiento incierto. Una sentencia como $P \vee Q$ nos permite expresar la incertidumbre acerca de cuál de los dos átomos es cierto, pero aún no hemos hablado de cómo podríamos representar *cuánta certeza* tenemos acerca de P, o de Q.

En la lógica clásica, podemos deducir Q a partir de P y $P \supset Q$; es decir, si un agente conoce $P \supset Q$, y posteriormente aprende P, entonces puede inferir Q. La pregunta es ¿existen procesos de inferencia análogos cuando la información es incierta? Se han empleado varios formalismos para representar y razonar con información incierta. Ya hemos hablado de algunos utilizados por MYCIN y PROSPECTOR. El formalismo que está más desarrollado (¡y alguien podría decir «el más apropiado!) es el basado en probabilidades. Comenzaremos el capítulo con un breve repaso de los fundamentos de la teoría de probabilidades.

Repaso a la teoría de probabilidades

19.1.1. Conceptos fundamentales

Asumimos que tenemos un conjunto de *variables aleatorias* V_1, V_2, ..., V_k. Cuando queremos hablar acerca del valor de V_i sin decir qué valor es, utilizamos el símbolo v_i. En nuestras aplicaciones, las variables aleatorias representan las características que nos interesan de un dominio. Los valores de las variables aleatorias pueden ser de diferentes tipos. Si las variables representan proposiciones, sus valores son *Verdadero* o *Falso* (o numéricamente, 1 o 0); si las variables representan medidas físicas (tales como altura, anchura, velocidad, etc.), los valores son numéricos; si las variables representan categorías (como color, letras del alfabeto, etc.), los valores son categóricos. Por ejemplo, el resultado de lanzar una moneda se podría representar por la variable *M*, cuyo valor podría ser uno de los dos valores categóricos *CA* (cara), o *CR* (cruz). Si estamos

hablando del resultado de lanzar k veces una moneda, necesitaríamos k variables (M_1, ..., M_k), cada una de las cuales podría tener el valor *CA* o *CR*.

Denotamos la *probabilidad conjunta* de que los valores V_1, V_2, ..., V_k sean v_1, v_2, ..., v_k respectivamente, con la expresión $p(V_1 = v_1, V_2 = v_2, ..., V_k = v_k)$. A dicha expresión se la denomina *función de probabilidad conjunta* sobre las variables V_1, V_2, ..., V_k. Esta función proyecta las variables de $p(V_1, V_2, ..., V_k)$ a los números reales en el intervalo entre 0 y 1. La sustitución de las variables de $p(V_1, V_2, ..., V_k)$ por valores concretos nos da la expresión $p(v_1, v_2, ..., v_k)$, que es una abreviación de $p(V_1 = v_1, V_2 = v_2, ..., V_k = v_k)$. De este modo, para un lanzamiento medio de moneda, podríamos tener $p(CA) = 1/2$; y si lanzamos una moneda cinco veces, podríamos tener $p(CA, CR, CR, CA, CR) = 1/32$, donde $p(CA, CR, CR, CA, CR)$ es la probabilidad conjunta de que el primer lanzamiento sea cara, el segundo cruz, el tercero cruz, el cuarto cara, y el quinto cruz.

Las funciones de probabilidad deben satisfacer ciertas propiedades; entre ellas, tenemos:

a) $0 \leq p(V_1, V_2, ..., V_k) \leq 1$

para cualquier asignación de las variables, y

b) $\sum p(V_1, V_2, ..., V_k) = 1$

que es el sumatorio de todas las asignaciones posibles de las variables. Así, en nuestro ejemplo del lanzamiento de monedas, que $p(CA) = 1/2$ es consistente con la propiedad (a); y como $p(CA) = 1/2$, al aplicar la propiedad (b), se restringe el valor de $p(CR)$ a 1/2. No tenemos mucho que decir acerca de cómo se asignan las probabilidades a los valores de las variables aleatorias. Igual que en la lógica proposicional, la verdad o falsedad de varias proposiciones, denotadas por fbfs, se basaba en el juicio subjetivo de los expertos en el dominio de aplicación (o mediante el procesamiento perceptual de los datos sensoriales); los valores de las probabilidades de las variables aleatorias, también dependerán del juicio del experto o del procesamiento perceptual. Nuestro principal interés es acerca de cómo realizar los cálculos que nos darán las probabilidades de ciertas variables que nos interesen.

En las aplicaciones que vamos a ver en este capítulo, las variables se corresponden con proposiciones acerca del dominio. Estas proposiciones pueden ser ciertas o falsas, por lo que las variables correspondientes tendrían los valores *Verdadero* o *Falso*. Podemos no tener certeza acerca del valor de verdad de una o más de estas proposiciones; esta incertidumbre estará representada por la probabilidad de los valores de las variables correspondientes. Por tanto, las técnicas que vamos a describir en este capítulo se pueden interpretar como las alternativas probabilísticas a los métodos expuestos en los Capítulos 13 y 14 acerca de la representación y el razonamiento mediante lógica proposicional (el desarrollo de alternativas probabilísticas a la lógica de primer orden sigue siendo, hoy en día, una frontera para la investigación; véase, por ejemplo, [Nilsson, 1986, y Glesner y Koller, 1995]).

Será de gran ayuda un ejemplo concreto para enmarcar nuestra introducción sobre los conceptos más importantes de la teoría de las probabilidades. Utilizaremos el mismo ejemplo que utilizamos en nuestra presentación sobre el razonamiento en el cálculo proposicional. Recordemos los átomos proposicionales `Bateria_OK`, `Se_mueve` y `Objeto_elevable`, que se definían para representar, respectivamente, que la batería estaba cargada, que el brazo del robot se movía (cuando sujetaba un bloque) y que un bloque se podía levantar. A estos átomos, añadiremos el átomo `Indicador`, que se crea para representar que el indicador que muestra el estado de la batería nos dice si la batería está completamente cargada. Para hacer nuestros diagramas y fórmulas menos pesados, renombraremos estos átomos con las letras `B`, `S`, `O` e `I`. Ahora supongamos que no estamos seguros acerca de si estos átomos tienen el valor *Verdadero* o *Falso*. Antes de que se pueda capturar cualquier lectura de los sensores, disponemos de una serie de probabilidades *a priori* sobre diversas combinaciones de los valores de las variables; es decir, por ejemplo, podemos pensar que es poco probable que `M` sea *Falso* cuando todas las demás variables son *Verdadero*.

Como tenemos cuatro variables binarias, tenemos 16 probabilidades sobre estas variables, cada una en la forma $p(\mathtt{B} = b, \mathtt{S} = s, \mathtt{O} = o, \mathtt{I} = i)$, donde b, s, o e i pueden ser *Verdadero* o *Falso*. El diseñador del agente debe especificar estos 16 valores, sujetos a las restricciones de que cada uno esté entre 0 y 1, y que el sumatorio de todos ellos debe ser 1. Como ejemplo, en la siguiente tabla listamos algunas de estas probabilidades conjuntas:

(B, S, O, I)	**Probabilidad conjunta**
Verdadero, Verdadero, Verdadero, Verdadero	0,5686
Verdadero, Vardadero, Verdadero, Falso	0,0299
Berdadero, Verdadero, Falso, Verdadero	0,0135
Verdadero, Verdadero, Falso, Falso	0,0007
...	

(Por supuesto, es muy difícil que el diseñador especifique las probabilidades con el grado de precisión que se muestra en la tabla. Nosotros lo hemos hecho para que estos valores sean consistentes con otras probabilidades, relacionadas con el ejemplo, que utilizaremos más adelante en este capítulo.)

Cuando conocemos los valores de todas las probabilidades conjuntas del conjunto de variables aleatorias, podemos calcular lo que denominamos *probabilidad marginal* de una variable aleatoria. Por ejemplo, la probabilidad marginal $p(\mathtt{B} = b)$ se define como el sumatorio de las 8 probabilidades conjuntas (la mitad de las 16) en las que $\mathtt{B} = b$:

$$p(\mathtt{B} = b) = \sum_{\mathtt{B}=b} p(\mathtt{B}, \mathtt{S}, \mathtt{O}, \mathtt{I})$$

Al utilizar esta fórmula, obtenemos la probabilidad marginal $p(\mathtt{B} = \textit{Verdadero}) = 0{,}95$, que es el sumatorio de las 8 probabilidades conjuntas en las que `B` es *Verdadero*.

Las probabilidades conjuntas de orden menor, también se pueden calcular mediante el sumatorio de las probabilidades conjuntas correspondientes. Por ejemplo, la probabilidad conjunta $p(\mathtt{B} = b, \mathtt{S} = s)$ es el sumatorio de las 4 probabilidades conjuntas en las que $\mathtt{B} = b$ y $\mathtt{S} = s$.

$$p(\mathtt{B} = b, \mathtt{S} = s) \sum_{\mathtt{B}=b, \mathtt{S}=s} p(\mathtt{B}, \mathtt{S}, \mathtt{O}, \mathtt{I})$$

De ello, se sigue que, cuando se conocen las probabilidades conjuntas de orden menor, podemos utilizar éstas para calcular otras probabilidades marginales, u otras probabilidades conjuntas de orden menor. De este modo, por ejemplo:

$$p(\mathtt{B} = b) = \sum_{\mathtt{B}=b} p(\mathtt{B}, \mathtt{S})$$

y

$$p(\mathtt{B} = b, \mathtt{S} = s) = \sum_{\mathtt{B}=b, \mathtt{S}=s} p(\mathtt{B}, \mathtt{S}, \mathtt{O})$$

Cuando tratamos con variables proposicionales (aquellas que tienen los valores *Verdadero* o *Falso*), a menudo empleamos una notación abreviada. Por ejemplo, algunas veces, en vez de

tener que escribir $p(\mathtt{B} = \textit{Verdadero}, \mathtt{S} = \textit{Falso})$, escribiremos $p(\mathtt{B}, \neg\mathtt{S})$, asumiendo que las variables afirmadas han sido instanciadas a *Verdadero* y las negadas a *Falso*. Esta notación sólo se utilizará cuando se deduzaca claramente del contexto que denotamos una probabilidad a las instanciaciones de las variables, más que una distribución de probabilidades sobre éstas.

Así, a partir de la función de probabilidad conjunta (en forma de tabla) para un conjunto de variables aleatorias, en principio, podremos calcular todas las probabilidades marginales y todas las probabilidades conjuntas de orden menor. Sin embargo, cuando tenemos un gran número de variables aleatorias, la tarea de especificar todas las probabilidades conjuntas, dejando aparte el cálculo de las probabilidades de orden menor, se hace intratable. Afortunadamente, en muchas aplicaciones, las probabilidades conjuntas satisfacen ciertas condiciones especiales que nos permiten la especificación y la realización de los cálculos de forma más viable. Más adelante describiremos estas condiciones.

19.1.2. Probabilidades condicionales

Ahora, lo que queremos es poder utilizar la información de los valores de algunas variables para obtener las probabilidades de otras. Por ejemplo, si el robot apilador de bloques percibe que su brazo no se mueve, podría querer calcular la probabilidad de que (dado ese hecho) la batería esté cargada. Estos cálculos se denominan *inferencias probabilísticas*, como analogía a los métodos de inferencia lógica. Antes de explicar cómo se pueden realizar las inferencias probabilísticas debemos definir lo que llamamos *probabilidades condicionales*.

La función de probabilidad condicional de V_i, dado V_j se denota por $p(V_i \mid V_j)$. Para cualquier valor de V_i y V_j, está función se define como:

$$p(V_i|V_j) = \frac{p(V_i, V_j)}{p(V_j)}$$

donde $p(V_i\ V_j)$ es la probabilidad conjunta de V_i y V_j, y $p(V_j)$ es la probabilidad marginal de V_j. A partir de esta expresión, podemos ver que también podemos escribir la probabilidad conjunta sobre la base de la probabilidad condicional:

$$p(V_i\ V_j) = p(V_i \mid V_j)p(V_j)$$

Volviendo al ejemplo del robot apilador de bloques, podemos calcular la probabilidad de que la batería esté cargada dado que el brazo del robot no se mueve mediante:

$$p(\mathtt{B} = \textit{Verdadero} \mid \mathtt{S} = \textit{Falso}) = \frac{p(\mathtt{B} = \textit{Verdadero}, \mathtt{S} = \textit{Falso})}{p(\mathtt{S} = \textit{Falso})}$$

Tanto el numerador como el denominador de esta expresión se pueden calcular a partir de los sumatorios de las probabilidades conjuntas, tal como ya hemos explicado.

Las probabilidades condicionales son fáciles de entender bajo una interpretación de las probabilidades como *frecuencias*. En esta interpretación, por ejemplo, $p(\mathtt{S} = \textit{Falso})$, es la relación (ratio) entre el número de veces que el brazo no se mueve y el total de veces que se intenta mover (en algún experimento imaginario llevado a cabo un número infinito de veces). Entonces, la probabilidad de que la batería esté cargada, dado que el brazo no se mueve, es el número de veces en que el brazo no se mueve y que la batería está cargada, en relación al número de veces en que el

brazo no se mueve. De este modo, una probabilidad condicional es la versión normalizada de una probabilidad conjunta.

Los diagramas de Venn [1], como el que se muestra en la Figura 19.1, nos son útiles para ilustrar los conceptos de probabilidad conjunta y condicional (para un número pequeño de variables). En este diagrama, mostramos dos elipses que se solapan, una denota las veces que el brazo no se mueve ($\mathtt{S} = \textit{Falso}$), y la otra denota las veces que la batería está cargada ($\mathtt{B} = \textit{Verdadero}$). Las áreas de las elipses, indicadas mediante notación abreviada en la figura, son proporcionales a sus probabilidades (marginales). El área exterior a ambas elipses se corresponde con las veces en las que el brazo se mueve y la batería no está cargada ($p(\mathtt{S} = \textit{Verdadero}, \mathtt{B} = \textit{Falso})$).

Fijémonos atentamente en las tres regiones disjuntas de las elipses, que son aquellas que se corresponden con las ocurrencias conjuntas de: el brazo no se mueve y la batería no está cargada; el brazo no se mueve y la batería está cargada y el brazo se mueve y la batería está cargada. Las áreas de cada una de estas regiones disjuntas son proporcionales a sus probabilidades conjuntas, tal como se muestra en el diagrama. La forma en que calculamos una probabilidad marginal a partir de las probabilidades conjuntas se saca fácilmente del diagrama: $p(\mathtt{B}) = p(\mathtt{B}, \mathtt{S}) + p(\mathtt{B}, \neg\mathtt{S})$.

También podemos tener probabilidades condicionales conjuntas de un conjunto de variables condicionadas por otro conjunto de variables. Por ejemplo (con notación abreviada):

$$p(\neg\mathtt{I}, \mathtt{B} \mid \neg\mathtt{S}, \mathtt{O}) = \frac{p(\neg\mathtt{I}, \mathtt{B}, \neg\mathtt{S}, \mathtt{O})}{p(\neg\mathtt{S}, \mathtt{O})}$$

Al calcular cualquier probabilidad condicional, las probabilidades conjunta y marginal que aparecen en el cómputo se pueden calcular a partir de cualquier conjunto completo de probabilidades conjuntas que contenga las variables requeridas, tal como hemos descrito antes.

También podemos expresar una probabilidad conjunta sobre la base de una *cadena* de probabilidades condicionales. Por ejemplo:

$$p(\mathtt{B}, \mathtt{O}, \mathtt{I}, \mathtt{S}) = p(\mathtt{B} \mid \mathtt{O}, \mathtt{I}, \mathtt{S})p(\mathtt{O} \mid \mathtt{I}, \mathtt{S})p(\mathtt{I} \mid \mathtt{S})p(\mathtt{S})$$

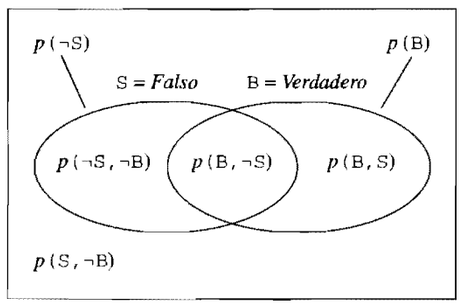


Figura 19.1.

Un diagrama de Venn.

[1] John Venn fue un lógico inglés [Venn, 1880].

La forma genérica de esta *regla encadenada* es:

$$p(V_1, V_2, \ldots, V_k) = \prod_{i=1}^{k} p(V_i | V_{i-1}, \ldots, V_1)$$

La expresión de la regla encadenada depende de la forma en que escojamos el orden de V_i. Distintas ordenaciones nos dan distintas expresiones, pero todas ellas nos deben dar el mismo resultado para el mismo conjunto de variables.

Como el orden en que estén las variables en una función de probabilidad conjunta no es importante (siempre que recordemos cuál es cuál), podemos escribir:

$$p(V_i, V_j) = p(V_i \mid V_j)p(V_j) = p(V_j \mid V_i)p(V_i) = p(V_j, V_i)$$

Fijémonos entonces en que:

$$p(V_i \mid V_j) = \frac{p(V_j|V_i)p(V_i)}{p(V_j)}$$

Esta última ecuación es muy importante. Se llama *teorema de Bayes*[2].

Vamos a comentar una notación final. Cuando tenemos las probabilidades conjuntas de un conjunto de variables, o las probabilidades condicionales de un conjunto de variables, será conveniente utilizar la notación de conjuntos. Así, alguna veces se utilizará $p(\mathcal{V})$ como abreviación de $p(V_1, ..., V_k)$, donde $\mathcal{V} = \{V_1, ..., V_k\}$. De manera similar, podríamos utilizar la abreviación $p(\mathcal{V}, \mathcal{V}_j)$, donde $\mathcal{V}_j$ también es un conjunto de variables. Si las variables $(V_1, ..., V_k)$ tienen los valores $v_1, ..., v_k$ respectivamente, lo denotaremos con la expresión $\mathcal{V} = \mathbf{v}$, donde $\mathcal{V}$ y $\mathbf{v}$ serán listas ordenadas.

Inferencia probabilística

19.2.1. Un método general

El escenario genérico para la inferencia probabilística es aquel en el que tenemos un conjunto $\mathcal{V}$, de variables proposicionales $V_1, ..., V_k$, en el que nos dan, en forma de evidencia, un subconjunto de variables de $\mathcal{V}$, al que llamaremos $\mathcal{E}$, con unos valores definidos $\mathcal{E} = \mathbf{e}$ (*Verdadero* o *Falso*). En las aplicaciones de agentes, las variables «disponibles» tendrán, por lo general, unos valores determinados por los procesos perceptuales. Lo que deseamos es, calcular la probabilidad condicional $p(V_i = v_i \mid \mathcal{E} = \mathbf{e})$, es decir, la probabilidad de que alguna variable V_i tenga el valor v_i, dada dicha evidencia. A este proceso lo llamaremos *inferencia probabilística*.

Como V_i puede tener los valores *Verdadero* o *Falso*, existen dos probabilidades condicionales en las que podríamos estar interesados, a saber, $p(V_i = Verdadero \mid \mathcal{E} = \mathbf{e})$ y $p(V_i = Falso \mid \mathcal{E} = \mathbf{e})$. Aun así, sólo necesitaremos calcular una de ellas, ya que $p(V_i = Verdadero \mid \mathcal{E} = \mathbf{e}) + p(V_i = Falso \mid \mathcal{E} = \mathbf{e}) = 1$, independientemente del valor de $\mathcal{E}$. Vamos a utilizar un método de cálculo

[2] El teorema de Bayes se formuló por primera vez por el reverendo Thomas Bayes [Bayes, 1763].

«exahustivo» para obtener $p(V_i = Verdadero \mid \mathcal{E} = \mathbf{e})$. Si utilizamos la definición de probabilidad condicional, tenemos:

$$p(V_i = Verdadero \mid \mathcal{E} = \mathbf{e}) = \frac{p(V_i = Verdadero, \mathcal{E} = \mathbf{e})}{p(\mathcal{E} = \mathbf{e})}$$

$p(V_i = Verdadero, \mathcal{E} = \mathbf{e})$ se obtiene utilizando la regla de cálculo de probabilidades de orden menor a partir de las de orden superior:

$$p(V_1 = Verdadero, \mathcal{E} = \mathbf{e}) = \sum_{V_i = Verdadero, \mathcal{E} = \mathbf{e}} p(V_1, \ldots V_k)$$

donde V_i, $i = 1, ..., k$ define nuestro conjunto de variables proposicionales. Es decir, calculamos el sumatorio de todos los valores de las probabilidades conjuntas de cada $V_i = Verdadero$, cuyos valores se obtienen de las variables de la evidencia. El cálculo de $p(\mathcal{E} = \mathbf{e})$ se puede realizar de forma similar, aunque, tal como se verá, no se necesita calcular de forma explícita.

Como ejemplo, supongamos que tenemos la probabilidades conjuntas:

$p(\mathrm{P}, \mathrm{Q}, \mathrm{R}) = 0{,}3$
$p(\mathrm{P}, \mathrm{Q}, \neg\mathrm{R}) = 0{,}2$
$p(\mathrm{P}, \neg\mathrm{Q}, \mathrm{R}) = 0{,}2$
$p(\mathrm{P}, \neg\mathrm{Q}, \neg\mathrm{R}) = 0{,}1$
$p(\neg\mathrm{P}, \mathrm{Q}, \mathrm{R}) = 0{,}05$
$p(\neg\mathrm{P}, \mathrm{Q}, \neg\mathrm{R}) = 0{,}1$
$p(\neg\mathrm{P}, \neg\mathrm{Q}, \mathrm{R}) = 0{,}05$
$p(\neg\mathrm{P}, \neg\mathrm{Q}, \neg\mathrm{R}) = 0{,}0$

Tenemos $\neg\mathrm{R}$ como evidencia y queremos calcular $p(\mathrm{Q} \mid \neg\mathrm{R})$. Utilizando el procedimiento anterior, calculamos:

$$p(\mathrm{Q}|\neg\mathrm{R}) = \frac{p(\mathrm{Q}, \neg\mathrm{R})}{p(\neg\mathrm{R})} = \frac{[p(\mathrm{P}, \mathrm{Q}, \neg\mathrm{R}) + p(\neg\mathrm{P}, \mathrm{Q}, \neg\mathrm{R})]}{p(\neg\mathrm{R})}$$
$$= \frac{(0{,}2 + 0{,}1)}{p(\neg\mathrm{R})} = \frac{0{,}3}{p(\neg\mathrm{R})}$$

Ahora bien, podemos calcular la probabilidad marginal $p(\neg\mathrm{R})$ directamente, o (lo que se suele hacer) calcular $p(\neg\mathrm{Q} \mid \neg\mathrm{R})$ con el mismo método, evitando el cálculo de $p(\neg\mathrm{R})$, aprovechando que $p(\mathrm{Q} \mid \neg\mathrm{R}) + p(\neg\mathrm{Q} \mid \neg\mathrm{R}) = 1$. Utilizaremos el segundo método:

$$p(\mathrm{Q}|\neg\mathrm{R}) = \frac{p(\neg\mathrm{Q}, \neg\mathrm{R})}{p(\neg\mathrm{R})} = \frac{[p(\mathrm{P}, \neg\mathrm{Q}, \neg\mathrm{R}) + p(\neg\mathrm{P}, \neg\mathrm{Q}, \neg\mathrm{R})]}{p(\neg\mathrm{R})}$$
$$= \frac{(0{,}1 + 0{,}0)}{p(\neg\mathrm{R})} = \frac{0{,}1}{p(\neg\mathrm{R})}$$

Ya que la suma de estos dos resultados debe ser igual a uno, tenemos que $p(\mathrm{Q} \mid \neg\mathrm{R}) = 0{,}75$.

Por lo general, la inferencia probabilística mediante este método es intratable, ya que para aplicarlo a casos en los que tenemos k variables, necesitamos un lista explícita de las 2^k probabilidades conjuntas $p(V_1, V_2, ..., V_k)$. Para muchos problemas que nos interesan, no podríamos anotar dicha lista aun conociéndola (que no suele ser el caso).

En vista a este problema, nos deberíamos preguntar «¿cómo razonan las personas con incertidumbre de forma eficiente?». Pearl [Pearl, 1986; Pearl 1988, y Pearl, 1990] conjeturó que lo hacemos formulando nuestro conocimiento sobre un dominio de una forma especial, que simplifica enormemente los cálculos de las probabilidades condicionales de ciertas variables, dada la evidencia, sobre de ellas. Esta forma de representar el conocimiento eficientemente involucra lo que se denominan *independencias condicionales* sobre las variables, el tema del que vamos a hablar seguidamente.

19.2.2. Independencia condicional

Decimos que una variable V es *condicionalmente independiente* de un conjunto de variables $\mathcal{V}_i$, dado otro conjunto $\mathcal{V}_j$, cuando $p(V \mid \mathcal{V}_i, \mathcal{V}_j) = p(V \mid \mathcal{V}_j)$. Utilizamos la expresión $I(V, \mathcal{V}_i \mid \mathcal{V}_j)$ para manifestar este hecho. La idea intuitiva es que si tenemos $I(V, \mathcal{V}_i \mid \mathcal{V}_j)$, entonces $\mathcal{V}_i$ no nos da más información sobre V de la que ya sabíamos con $\mathcal{V}_j$, y en lo que se refiere a V, si conocemos $\mathcal{V}_j$, podemos ignorar $\mathcal{V}_i$. En nuestro ejemplo del apilamiento de bloques, parace razonable pensar que si ya sabemos (mediante cualquier otro medio) que la batería está cargada (B = *Verdadero*) entonces, en tanto que estemos interesados acerca de si el brazo se mueve (S), no necesitamos el conocimiento explícito acerca de I (el indicador que carga de la batería), es decir, $p(\mathtt{S} \mid \mathtt{B}, \mathtt{I}) = p(\mathtt{S} \mid \mathtt{B})$.

Si una variable V_i es condicionalmente independiente de otra variable V_j, dado un conjunto $\mathcal{V}$, tenemos (por definición) que $p(V_i|V_j, \mathcal{V}) = p(V_i|\mathcal{V})$. A partir de la definición de la probabilidad condicional, tenemos que $p(V_i|V_j, \mathcal{V})p(V_j|\mathcal{V}) = p(V_i, V_j|\mathcal{V})$. Combinando estas dos expersiones, obtenemos:

$$p(V_i, V_j|\mathcal{V}) = p(V_i|\mathcal{V})p(V_j|\mathcal{V})$$

para el caso en el que $I(V_i, V_j|\mathcal{V})$. Nos debemos fijar en que V_i y V_j aparecen de forma simétrica. Por tanto, decir que V_i es condicionalmente independiente de V_j, dado $\mathcal{V}$, es lo mismo que decir que V_j es condicionalmente independiente de V_i, dado $\mathcal{V}$. Esto es suficiente para decir que V_i y V_j son condicionalmente independientes dado $\mathcal{V}$. Este mismo resultado se puede aplicar a conjuntos de variables, a saber, si $\mathcal{V}_i$ y $\mathcal{V}_j$ son condicionalmente independientes dado $\mathcal{V}$ (es decir, $I(\mathcal{V}_i, \mathcal{V}_j|\mathcal{V})$), entonces $p(\mathcal{V}_i, \mathcal{V}_j|\mathcal{V}) = p(\mathcal{V}_i|\mathcal{V})p(\mathcal{V}_j|\mathcal{V})$. Si $\mathcal{V}$ está vacío ($\varnothing$), simplemente decimos que $\mathcal{V}_i$ y $\mathcal{V}_j$ son independientes.

Como generalizacion de la independencia entre parejas, decimos que las variables $V_1, ..., V_k$ son *mutuamente independientes*, dado un conjunto $\mathcal{V}$, si cada una de las variables es condicionalmente independiente de todas las demás, dado $\mathcal{V}$. Como:

$$p(V_1, V_2, ..., V_k|\mathcal{V}) = \prod_{i=1}^{k} p(V_i|V_{i-1}, ..., V_1, \mathcal{V})$$

y, como cada V_i es condicionalmente independiente de las otras dado $\mathcal{V}$, tenemos que:

$$p(V_1, V_2, ..., V_k|\mathcal{V}) = \prod_{i=1}^{k} p(V_i|\mathcal{V})$$

En el caso en que $\mathcal{V}$ esté vacío, tenemos:

$$p(V_1, V_2, ..., V_k) = p(V_1)p(V_2) ... p(V_k)$$

y decimos que las variables son *incondicionalmente independientes*.

Las independencias condicionales se pueden representar adecuadamente con las estructuras denominadas *redes bayesianas* (también se las denomina *redes de creencias*). Estas estructuras son muy útiles en la inferencia probabilística. Las independencias condicionales representadas mediante redes bayesianas nos llevan a grandes ahorros en los cálculos de la inferencia probabilística.

Redes bayesianas

Una red bayesiana es un grafo dirigido acíclico (GDA) cuyos nodos están etiquetados con variables aleatorias. Una red bayesiana estipula que cada nodo V_i del grafo es condicionalmente independiente de cualquier subconjunto de nodos que no sean descendientes de V_i, dados sus padres, es decir, $\mathcal{A}(V_i)$ es *cualquier* subconjunto del grafo que no es descendiente de V_i, y $\mathcal{P}(V_i)$ son los padres de V_i en el grafo. El grafo sólo es una forma de representar que, para todo V_i en el grafo, $I(V_i, \mathcal{A}(V_i)|\mathcal{P}(V_i))$, es lo mismo que decir que $p(V_i|\mathcal{A}(V_i), \mathcal{P}(V_i)) = p(V_i|\mathcal{P}(V_i))$.

Sean $V_1, V_2 ..., V_k$ los nodos de una red bayesiana. Con la asunción de independencia condicional que hemos descrito para la red, podemos definir la probabilidad conjunta de todos los nodos como:

$$p(V_1, V_2, ..., V_k) = \prod_{i=1}^{k} p(V_i|\mathcal{P}(V_i))$$

Esta expresión se puede derivar fácilmente aplicando las independencias condicionales a la expresión de la regla encadenada de la probabilidad conjunta de todas las variables utilizando cualquier orden de la regla que sea consistente con el orden parcial implícito en la red bayesiana (un GDA).

Algunas veces, a las redes bayesianas se las denomina *redes causales* cuando los arcos que conectan los nodos se pueden interpretar como la representación de relaciones causales directas. A menudo, los expertos son capaces de relacionar causas y efectos de una forma que revelan independencias condicionales inherentes que son representables mediante una red bayesiana. La estructuración de redes bayesianas utilizando la noción intuitiva de causalidad, por lo general, genera redes para las cuales las asunciones de independencia condicional resultan ser adecuadas. En las palabras del investigador [Heckerman, 1996, p. 14]: «... para construir una red bayesiana de un conjunto de variables dibujamos arcos desde las variables (nodos) causa a los nodos efecto. En la mayoría de los casos, se obtiene una red causal [cuyas implicaciones de independencia condicional suelen ser precisas]».

Vamos a ilustrar la construcción de una red bayesiana mediante nuestro ejemplo del apilamiento de bloques. Comenzaremos con lo que imaginamos son «las primeras causas» en este dominio, a saber, las variables que se corresponden con las proposiciones «la batería está cargada» (B) y «el bloque se puede levantar» (L). B y L son causas de S («el brazo se mueve»), y B es una causa de I («el indicador muestra que la batería está cargada»). Así, dibujaríamos la red bayesiana de este problema como la que se muestra en la Figura 19.2. Fijémonos en que, sobre otras cosas, la red explicita que $p(\mathtt{S}|\mathtt{I}, \mathtt{B}, \mathtt{L}) = p(\mathtt{S}|\mathtt{B}, \mathtt{L})$. Si hubiese un nodo en la red, llamémoslo

E (cuyo significado fuese que el bloque está elevado), no podría suceder que $p(S|Y, B, L, E) = p(S|B, L)$, porque E sería un descendiente de S (que el bloque esté elevado influye a la probabilidad de que el brazo se mueva. ¿De qué otra forma podría elevarse el bloque?). La expresión de la probabilidad conjunta de todos los nodos de la red se muestra en la figura.

Vemos que para poder calcular el valor de las probabilidades conjuntas dadas por la red bayesiana necesitamos conocer las funciones de probabilidad condicional de cada nodo de la red que esté influido por sus padres, tal como se puede observar en la Figura 19.2. Para los nodos sin padres, las probabilidades no están condicionadas por otros nodos, por lo que se les denomina *probabilidades a priori* de estas variables (nodos). Por tanto, una especificación completa de las probabilidades de un conjunto de variables aleatorias significa tener una red bayesiana de estas variables, junto a las tablas de *probabilidades condicionales* (TPC) de cada nodo de la red.

La fórmula de la función de probabilidad conjunta de la red bayesiana de nuestro ejemplo del apilamiento de bloques se podría comparar con una parecida (sin independencias condicionales), que se obtendría mediante la regla encadenada:

$$p(I, S, B, L) = p(I|B, S, L)p(S|B, L)p(B|L)p(L)$$

Fijémonos en que la fórmula de la red bayesiana es mucho más sencilla. Sin las independencias condicionales especificadas en la red bayesiana, la especificación de la probabilidad conjunta para las cuatro variables del ejemplo requieren la especificación de 16 probabilidades conjuntas distintas (en realidad, sólo de 15, ya que todas deben sumar 1). Tal como se evidencia en la Figura 19.2, las asunciones hechas en la red bayesiana nos permiten tener que calcular sólo ocho probabilidades. Cuando hay muchas independencias condicionales entre las variables del dominio, la expresión de la probabilidad conjunta que se calcula a partir de la red bayesiana necesita la especificación de muchas menos probabilidades que las que necesitaría sin dichas independencias. Esta reducción convierte en tratables muchos problemas que, de otro modo, no lo serían.

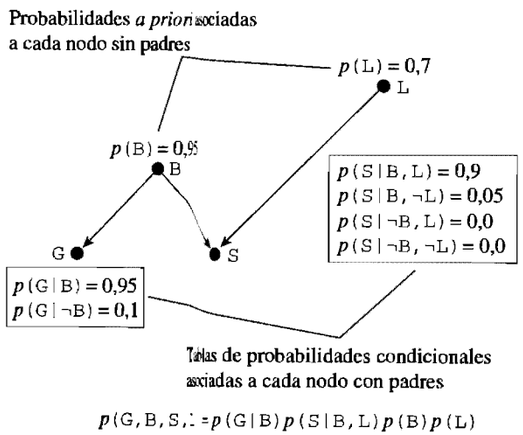


Figura 19.2.

Una red bayesiana.

Patrones de inferencia en redes bayesianas

Hay tres patrones importantes de inferencia en las redes bayesianas. Para explicarlos, seguiremos con nuestro ejemplo.

- Inferencia causal (o descendente). Supongamos que deseamos calcular $p(S|O)$, la probabilidad de que el brazo se mueva dado que el objeto se puede elevar. Como el hecho de que el objeto sea elevable es una de las causas de que el brazo pueda moverse, decimos que este cálculo es un ejemplo de *razonamiento causal*. A O se le denomina *evidencia utilizada* en la inferencia, y a S el *nodo de consulta*. Así es cómo realizamos la inferencia: primero, expandimos $p(S|O)$ (la probabilidad marginal) en el sumatorio de dos probabilidades conjuntas (porque debemos tener en cuenta al otro padre de S, B):

  $$p(S|O) = p(S, B|O) + p(S, \neg B|O)$$

  Luego, queremos que S esté condicionado, tanto a este otro padre, como a O, así que utilizamos la expresión de la regla encadenada para escribir:

  $$p(S|O) = p(S|B, O)p(B|O) + p(S|\neg B, O)p(\neg B|O)$$

  Pero $p(B|O) = p(B)$ (se obtiene de la propia estructura de la red; fijémonos en que B no tiene padres). E igualmente, $p(\neg B|O) = p(\neg B)$.

  Por tanto, $p(S|O) = p(S|B, O)p(B) + p(S|\neg B, O)p(\neg B)$. Como todas estas cantidades se dan en la red, podemos calcularlas y obtener:

  $$p(S|O) = 0,855$$

  Las operaciones que hemos realizado en este ejemplo merecen anotarse, porque pueden generalizarse a versiones más complejas de razonamiento causal (tal como veremos más adelante). Las principales operaciones son:

  a) Escribir las probabilidades condicionales del nodo de consulta *V*, dada la evidencia, en términos de la probabilidad conjunta de *V* y todos sus padres (los cuales no son la evidencia), dada la evidencia.
  b) Expresar esta probabilidad conjunta como la probabilidad de *V*, condicionada por todos sus padres.

- Inferencia de diagnóstico (o ascendente). Ahora vamos a calcular $p(\neg O|\neg S)$, la probabilidad de que el bloque no se pueda elevar dado que el brazo no se mueve. Aquí, los papeles de la consulta y de la evidencia son a la inversa a los del último ejemplo. Como estamos utilizando un efecto (o síntoma) para inferir una causa, a este tipo de razonamiento lo denominaremos *razonamiento de diagnóstico*.

  $$p(\neg O|\neg S) = \frac{p(\neg S|\neg O)p(\neg O)}{p(\neg S)} \text{ (teorema de Bayes)}$$

Ahora calculamos $p(\neg S|\neg O) = 0{,}9525$ (mediante razonamiento causal), y luego sustituimos en $p(\neg O|\neg S) = \frac{0{,}9525 \times 0{,}3}{p(\neg S)} = \frac{0{,}28575}{p(\neg S)}$. De igual modo, $p(L|\neg S) = \frac{p(\neg S \mid O)p(O)}{p(\neg S)} = \frac{0{,}0597 \times 0{,}7}{p(\neg S)} = \frac{0{,}03665}{p(\neg S)}$.

Y como ambas expresiones deben sumar 1, tenemos $p(\neg O|\neg S) = 0{,}88632$.

Los cálculos que hemos realizado en este sencillo ejemplo de razonamiento de diagnóstico también se pueden generalizar. El paso principal es la utilización del teorema de Bayes para convertir el problema a uno de razonamiento causal.

- Inferencia intercausal (o justificación). Si nuestra única evidencia es que $\neg S$ (el brazo no se mueve), podemos calcular la probabilidad de que el bloque no se pueda elevar, $\neg O$, tal como hicimos. Pero si también tenemos $\neg B$ (la batería no está cargada), entonces $\neg O$ debería ser menos cierto (tener menos peso). En este caso, decimos que $\neg B$ *justifica* $\neg S$, haciendo a $\neg O$ menos cierto. Este tipo de inferencia utiliza un proceso de razonamiento causal (o descendente) incrustado en un proceso de diagnóstico (o ascendente).

$$p(\neg O|\neg B, \neg S) = \frac{p(\neg S, \neg B \mid \neg O)p(\neg O)}{p(\neg B, \neg S)} \text{ (teorema de Bayes)}$$

$$= \frac{p(\neg S \mid \neg B, \neg O)p(\neg B \mid \neg O)p(\neg O)}{p(\neg B, \neg S)} \text{ (def. de la probabilidad condicional)}$$

$$= \frac{p(\neg S \mid \neg B, \neg O)p(\neg B)p(\neg O)}{p(\neg B, \neg S)} \text{ (estructura de la red bayesiana)}$$

De esta expresión, utilizando las probabilidades de la red, y resolviendo $p(\neg B, \neg S)$ de la forma habitual, obtenemos $p(\neg O|\neg B, \neg S) = 0{,}030$, que es tal como esperábamos, mucho menor que $p(\neg O|\neg S)$ (calculado anteriormente). Otra vez, nos debemos fijar en el uso del teorema de Bayes, que es un paso bastante importante de la justificación.

Evidencia con incertidumbre

La expresión $p(V|\varepsilon)$, donde V es un nodo consulta, no nos da la probabilidad correcta cuando la evidencia ε es incierta. En los cálculos sobre la red bayesiana, para poder «obtener» los nodos de evidencia, debemos tener la certeza sobre la verdad o la falsedad de las proposiciones que representan. Podemos alcanzar este requisitos mediante una distribución en la que tengamos cada nodo de evidencia (los que son inciertos) con un nodo hijo, del cual tengamos certeza. Entonces, en el último ejemplo que hemos visto (el de justificación) podríamos suponer que el robot no tiene certeza acerca de si su brazo no se está moviendo (podría tener un sensor poco fidedigno). En ese caso, la evidencia se podría obtener de un nodo S', que representaría la proposición «el sensor del brazo dice que el brazo se ha movido». Podemos tener la certeza acerca de si dicha proposición es cierta o falsa dependiendo de su lectura. Entonces, la red bayesiana se utilizaría para calcular $p(\neg O|\neg B, \neg S')$, en vez de $p(\neg O|\neg B, \neg S)$. Desde luego, las red necesitaría los valores de $p(S'|S)$ y de $p(S'|\neg S)$, que representarían la fiabilidad del sensor.

Fijémonos en que la red de la Figura 19.2 ya nos daba información acerca de que no podemos tener certeza sobre si la batería estaba o no cargada. El nodo B tiene un nodo hijo I, expresando cómo de fiable era el indicador mediante las probabilidades $p(I|B)$ y $p(I|\neg B)$. Dejamos al lector el cálculo de $p(\neg O|\neg B, \neg S')$ (quizá tras una lectura más avanzada).

Aun con las simplificaciones que nos da la red bayesiana, el método exhaustivo que acabamos de utilizar para el cálculo de las diversas probabilidades condicionales a partir de la probabilidad conjunta es, por lo general, intratable en redes grandes. En el peor caso, la complejidad del algoritmo crece de forma exponencial en relación al número de variables proposicionales. Por suerte, existen varios métodos abreviados para el cálculo de las probabilidades condicionales en redes con formas especiales. Consideraremos algunos de estos métodos después de presentar otra consecuencia de las independencias condicionales en las redes bayesianas.

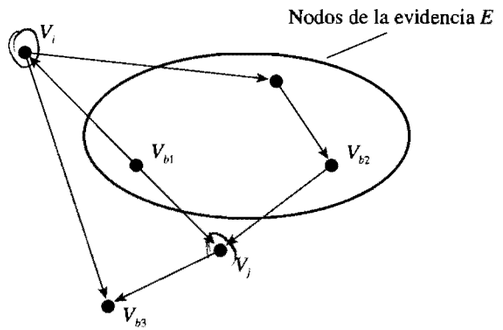
Separación-d

Ocurre que una red bayesiana tiene más independencias condicionales que las que sólo involucran a los padres de un nodo. Por ejemplo, en la Figura 19.2, $p(S|I, B) = p(S|B)$, es decir, S es condicionalmente independiente de I dado B (aunque no nos dan ambos padres de S). De forma intuitiva, el conocimiento del (efecto) I puede influir sobre el conocimiento de la (causa) B, el cual influye sobre el conocimiento del (otro efecto) S. Pero si nos dan la causa B, no hay nada que nos pueda decir I acerca de S. En este caso, decimos que B *separa-d* (separación dependiente de la dirección) a S de I.

Existen otros tipos de independencias condicionales en las redes bayesianas. Vamos a hablar de ellas, pero consúltese [Pearl, 1988, pp. 117-122] para ver las demostraciones.

Dos nodos V_i y V_j son independientes condicionalmente dado un conjunto de nodos ε (es decir, $I(V_i, V_j|\varepsilon)$) si por cada camino no dirigido entre V_i y V_j hay algún nodo V_b que cumple algunas de las siguientes tres propiedades (véase Fig. 19.3):

1. V_b pertenece a ε, y ambos arcos salen de V_b.
2. V_b pertenece a ε, y un arco va hacia V_b y el otro arco sale de él.
3. Ni V_b ni ningún descendiente suyo pertenecen a ε, y ambos arcos van hacia V_b.



Dado el conjunto de nodos de la evidencia, V_i es independiente de V_j debido a que los tres caminos se bloquean entre sí. Los nodos que bloquean son:

a) V_{b1} es un nodo evidencia, y ambos arcos salen de V_{b1}.
b) V_{b2} es un nodo evidencia, y un arco va hacia él y otro sale de él.
c) V_{b3} no es un nodo de la evidencia, ni ninguno de sus descendientes, y los dos arcos van a él.

Figura 19.3.

Independencia condicional mediante nodos bloqueadores.

Cuando cualquiera de estas tres condiciones se cumple, decimos que el nodo V_b *bloquea* el camino dado ε. Tengamos en cuenta que los caminos (rutas) de los que hablamos son caminos indirectos, es decir, caminos que ignoran la dirección de los arcos. Si *todos* los caminos entre V_i y V_j están bloqueados, entonces decimos que ε separa-d V_i de V_j (separación dependiente de la dirección), y concluimos que V_i y V_j son independientes condicionalmente dado ε. Ejemplos en la Figura 19.2 de independencia condicional por causa de la separación-d son:

- I(I, O|B); por la regla 1, B bloquea el (único) camino entre I y O, dado B; por la regla 3, S también bloquea este camino dado B, porque S no es miembro del conjunto de evidencia.
- I(I, O) y I(B, O); por la regla 3, S bloquea el (único) camino entre I y O, y entre B y O, dado el conjunto de evidencia vacío (S no es miembro del conjunto de evidencia).

Sin embargo, nos debemos fijar en que B y O no son condicionalmente independientes dado S, ya que, aunque S está en el camino entre B y O, los dos arcos de este camino van a S, y S pertenece al conjunto de evidencia; por tanto, S no bloquea el camino entre B y O.

El concepto de separación-d también se puede aplicar a conjuntos de nodos. Dos conjuntos $\mathcal{V}_i$ y $\mathcal{V}_j$ son condicionalmente independientes dado ε, si están separados-d por ε; es decir, si cada camino (ruta) no dirigido entre todos los nodos de $\mathcal{V}_i$ y todos los nodos de $\mathcal{V}_j$ están bloqueados dado ε.

Aun con la utilización de la separación-d, la inferencia probabilística en redes bayesianas es, por lo general, NP-duro [Cooper, 1990]. Sin embargo, se pueden hacer algunas simplificaciones para una clase muy importante de redes, denominada poliárboles. Un *poliárbol* es un GDA en el que, para cada par de nodos, sólo hay un camino a lo largo de los arcos en cualquier dirección (red de conexión única). Por ejemplo, la red de la Figura 19.2 es un poliárbol. Vamos a mostrar el funcionamiento de la inferencia probabilística en poliárboles con un ejemplo más desarrollado (el método que vamos a mostrar está basado en un algoritmo propuesto por [Russel y Norvig, 1995, p. 447 y ss.]).

Inferencia probabilística en poliárboles

La red que se muestra en la Figura 19.4 es el típico ejemplo de poliárbol. Lo que queremos es calcular la probabilidad de Q a partir de los otros nodos en la red. Fijémonos en que algunos nodos están conectados a Q sólo mediante sus padres; entonces diremos que estos nodos están *por encima* de Q. Los otros nodos están conectados sólo mediante sus sucesores inmediatos (sus hijos); diremos que éstos están *por debajo* de Q. También nos debemos fijar en que no hay ningún otro camino (ruta) excepto los que unen un nodo por encima de Q a un nodo por debajo de Q (si no la red ya no sería un poliárbol). ¡Estas definiciones y propiedades de conectividad se aplican a todos los nodos del poliárbol! Tendremos tres tipos de evidencia:

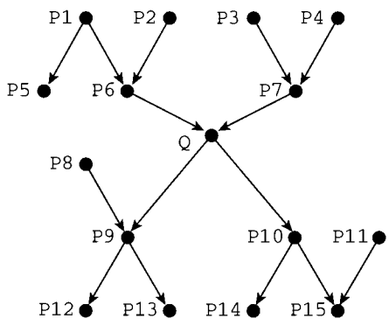


Figura 19.4.

Un poliárbol típico.

1. Todos los nodos de la evidencia están por encima de Q (apoyo causal). Como ejemplo típico de este tipo de evidencia, calcularemos p(Q|P5, P4).
2. Todos los nodos de la evidencia están por debajo de Q (apoyo evidencial). Como ejemplo típico de este tipo de evidencia, calcularemos p(Q|P12, P13, P14, P11).
3. Hay nodos de la evidencia por encima y por debajo de Q.

19.7.1. Apoyo causal

Vamos a calcular p(Q|P5, P4), donde todos los nodos de la evidencia están por encima de Q. El cálculo es un traceado de la ejecución de un algoritmo recursivo «ascendiente» que calcula la probabilidad de cada antecesor de Q, dada la evidencia, hasta que se alcanza la evidencia o hasta que la evidencia está por debajo del antecesor. El algoritmo se comporta de la siguiente manera:

Primero «involucramos a los padres» (de Q):

$$p(\text{Q}|\text{P5}, \text{P4}) = \sum_{\text{P6},\text{P7}} p(\text{Q}, \text{P6}, \text{P7}|\text{P5}, \text{P4})$$

(La notación especial que se utiliza para indexar este sumatorio es para indicar que añadimos cuatro versiones de p(Q, P6, P7|P5, P4), la original, una con ¬P6 en vez de P6, otra con ¬P7 en vez de P7, y la última con ¬P6 y ¬P7).

Luego, introducimos a los padres de Q como parte de la evidencia usando la definición de la independencia condicional:

$$p(\text{Q}, \text{P6}, \text{P7}|\text{P5}, \text{P4}) = p(\text{Q}|\text{P6}, \text{P7}, \text{P5}, \text{P4})p(\text{P6}, \text{P7}|\text{P5}, \text{P4})$$

Con la sustitución, obtenemos:

$$p(\text{Q}|\text{P5}, \text{P4}) = \sum_{\text{P6},\text{P7}} p(\text{Q}|\text{P6}, \text{P7}, \text{P5}, \text{P4})p(\text{P6}, \text{P7}|\text{P5}, \text{P4})$$

Ahora, como un nodo es condicionalmente independiente de sus antecesores, dados sus padres:

$$p(\text{Q}|\text{P5}, \text{P4}) = \sum_{\text{P6},\text{P7}} p(\text{Q}|\text{P6}, \text{P7})p(\text{P6}, \text{P7}|\text{P5}, \text{P4})$$

Entonces, la separación-d nos permite fragmentar a los padres:

$$p(\text{Q}|\text{P5}, \text{P4}) = \sum_{\text{P6},\text{P7}} p(\text{Q}|\text{P6}, \text{P7})p(\text{P6}|\text{P5}, \text{P4})p(\text{P7}|\text{P5}, \text{P4})$$

Y, finalmente, la separación-d nos permite ignorar la evidencia por encima de unos de los padres al calcular la probabilidad del otro padre:

$$p(\text{Q}|\text{P5}, \text{P4}) = \sum_{\text{P6},\text{P7}} p(\text{Q}|\text{P6}, \text{P7})p(\text{P6}|\text{P5})p(\text{P7}|\text{P4})$$

Es muy importante fijarnos en que los términos que se suman son *a*) las probabilidades del nodo de consulta, dados los valores de sus padres (cuyas probabilidades se dan en la red bayesiana), y *b*) las probabilidades de cada uno de los padres se dan a partir de la evidencia por encima de ellos (mediante una llamada recursiva al algoritmo que estamos describiendo). Estos resultados se siguen de forma directa por el hecho de que estamos trabajando con un poliárbol.

Este mismo procedimiento se aplica recursivamente hasta que finalmente se alcanzan los nodos que tienen un nodo de la evidencia como padre o los que no tienen padres (nodos que ellos mismos no son evidencia). El primero de los dos casos lo tenemos en el cálculo de $p(\mathtt{P7}|\mathtt{P4})$, donde el nodo de la evidencia P4 es uno de los padres del nodo de consulta P7. En este caso, el proceso de «involucrar a los padres» es más sencillo, ya que uno de los dos padres de P7 ya ha sido involucrado. Como $p(\mathtt{P7}, \mathtt{P3}|\mathtt{P4}) = p(\mathtt{P7}|\mathtt{P3}, \mathtt{P4})p(\mathtt{P3}|\mathtt{P4})$, podemos escribir:

$$p(\mathtt{P7}|\mathtt{P4}) = \sum_{\mathtt{P3}} p(\mathtt{P7}|\mathtt{P3}, \mathtt{P4})p(\mathtt{P3}|\mathtt{P4}) = \sum_{\mathtt{P3}} p(\mathtt{P7}|\mathtt{P3}, \mathtt{P4})p(\mathtt{P3})$$

(El último paso se sigue de $I(\mathtt{P3}, \mathtt{P4})$.) Todos los términos del sumatorio se obtienen de la red bayesiana, entonces, el procedimiento para el ejemplo finaliza en esta ramificación.

Al calcular $p(\mathtt{P6}|\mathtt{P5})$, obtenemos:

$$p(\mathtt{P6}|\mathtt{P5}) = \sum_{\mathtt{P1},\mathtt{P2}} p(\mathtt{P6}|\mathtt{P1}, \mathtt{P2})p(\mathtt{P1}|\mathtt{P5})p(\mathtt{P2})$$

Aquí debemos calcular $p(\mathtt{P1}|\mathtt{P5})$, y nos fijamos en que el nodo de la evidencia no se encuentra «por encima» del nodo de consulta, sino «por debajo», por lo que no podemos proseguir con la llamada recursiva, pero debemos utilizar el procedimiento de «apoyo evidencial» (que aún no hemos descrito). En este ejemplo, simplemente, utilizamos el teorema de Bayes para obtener $p(\mathtt{P1}|\mathtt{P5}) = \dfrac{p(\mathtt{P5}|\mathtt{P1})p(\mathtt{P1})}{p(\mathtt{P5})}$. Después, todas las cantidades que se necesitan para calcular $p(\mathtt{P6}|\mathtt{P5})$ se dan en la red bayesiana. Podemos juntar todos estos resultados (realizando todos los sumatorios) para obtener la respuesta final de $p(\mathtt{Q}|\mathtt{P5}, \mathtt{P4})$.

19.7.2. Apoyo evidencial

Seguidamente, vamos a calcular $p(\mathtt{Q}|\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11})$, donde todos los nodos evidencia están por debajo de Q. Nuestros cálculos, otra vez, van a ser el traceado de la ejecución de un algoritmo recursivo. Su funcionamiento es que en el nivel principal utilizamos la regla de bayes para escribir:

$$\begin{aligned} p(\mathtt{Q}|\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11}) &= \frac{p(\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11}|\mathtt{Q})p(\mathtt{Q})}{p(\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11})} \\ &= kp(\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11}|\mathtt{Q})p(\mathtt{Q}) \end{aligned}$$

donde $k = \dfrac{1}{p(\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11})}$ es un factor de normalización que se calculará más adelante, de la misma forma como lo hicimos en los ejemplos iniciales. Por separación-d $I(\{\mathtt{P12}, \mathtt{P13}\}, \{\mathtt{P14}, \mathtt{P11}\}|\mathtt{Q})$, obtenemos

$$p(\mathtt{Q}|\mathtt{P12}, \mathtt{P13}, \mathtt{P14}, \mathtt{P11}) = kp(\mathtt{P12}, \mathtt{P13}|\mathtt{Q})p(\mathtt{P14}, \mathtt{P11}|\mathtt{Q})p(\mathtt{Q})$$

Nos tenemos que fijar en que hemos separado el conjunto {P12, P13, P14, P11} en dos subconjuntos correspondientes a los dos hijos de Q. Cada uno de los términos $p(\mathtt{P12}, \mathtt{P13}|\mathtt{Q})$ y $p(\mathtt{P14}, \mathtt{P11}|\mathtt{Q})$ son un caso del cálculo de la probabilidad de un conjunto de nodos de consulta, dado un único nodo de la evidencia que está por encima de ellos. Entonces, podemos utilizar un algoritmo parecido al anterior. Como sólo hay un único nodo de la evidencia, es conveniente utilizar el algoritmo recursivo *descendente* en vez del *ascendente*.

Vamos a ilustrar el funcionamiento de la versión descendente en el cálculo de $p(\mathtt{P12}, \mathtt{P13}|\mathtt{Q})$. El paso clave es involucrar al único hijo de Q, P9, que está por encima del conjunto de nodos de consulta {P12, P13}. Primero, fijémonos en que $p(\mathtt{P12}, \mathtt{P13}, \mathtt{P9}|\mathtt{Q}) = p(\mathtt{P12}, \mathtt{P13}|\mathtt{P9}, \mathtt{Q})p(\mathtt{P9}|\mathtt{Q})$, por la definición de la independencia condicional. Entonces:

$$p(\mathtt{P12}, \mathtt{P13}|\mathtt{Q}) = \sum_{\mathtt{P9}} p(\mathtt{P12}, \mathtt{P13}|\mathtt{P9}, \mathtt{Q})p(\mathtt{P9}|\mathtt{Q})$$

Ahora, por la separación-d de $I(\{\mathtt{P12}, \mathtt{P13}\}, \mathtt{Q}|\mathtt{P9})$, tenemos

$$p(\mathtt{P12}, \mathtt{P13}|\mathtt{Q}) = \sum_{\mathtt{P9}} p(\mathtt{P12}, \mathtt{P13}|\mathtt{P9})p(\mathtt{P9}|\mathtt{Q})$$

De los términos del sumatorio, $p(\mathtt{P9}|\mathtt{Q})$ se calcula involucrando a todos los padres de P9:

$$p(\mathtt{P9}|\mathtt{Q}) = \sum_{\mathtt{P8}} p(\mathtt{P9}|\mathtt{P8}, \mathtt{Q})p(\mathtt{P8})$$

$p(\mathtt{P9}|\mathtt{P8}, \mathtt{Q})$ se da en la red. El otro término, $p(\mathtt{P12}, \mathtt{P13}|\mathtt{P9})$ es una llamada recursiva al procedimiento ascendente que calcula la probabilidad del conjunto de nodos de consulta dado un único nodo de la evidencia, que está por encima de ellos. En este caso, la llamada recursiva finaliza después de un paso, ya que los hijos de P9 son los nodos de la evidencia. Como P12 y P13 son independientes dado P9, tenemos que $p(\mathtt{P12}, \mathtt{P13}|\mathtt{P9}) = p(\mathtt{P12}, \mathtt{P9})p(\mathtt{P13}|\mathtt{P9})$. Ambas probabilidades se dan en la red.

Aplicando el procedimiento descendente sobre $p(\mathtt{P14}, \mathtt{P11}|\mathtt{Q})$, obtenemos:

$$p(\mathtt{P14}, \mathtt{P11}|\mathtt{Q}) = \sum_{\mathtt{P10}} p(\mathtt{P14}, \mathtt{P11}|\mathtt{P10})p(\mathtt{P10}|\mathtt{Q})$$

Y entonces:

$$p(\mathtt{P14}, \mathtt{P11}|\mathtt{Q}) = \sum_{\mathtt{P10}} p(\mathtt{P14}, \mathtt{P10}), p(\mathtt{P11}|\mathtt{P10})p(\mathtt{P10}|\mathtt{Q})$$

porque $I(\mathtt{P14}, \mathtt{P11}|\mathtt{P10})$. Sólo el término central de la multiplicación no se da en la red; entonces utilizamos el procedimiento descendente para calcularlo:

$$p(\mathtt{P11}|\mathtt{P10}) = \sum_{\mathtt{P15}} p(\mathtt{P11}|\mathtt{P15}, \mathtt{P10})p(\mathtt{P15}|\mathtt{P10})$$

Aquí:

$$p(\mathtt{P15}, \mathtt{P10}) = \sum_{\mathtt{P11}} p(\mathtt{P15}|\mathtt{P10}, \mathtt{P11})p(\mathtt{P11})$$

Pero en $p(\mathrm{P11}|\mathrm{P15}, \mathrm{P10})$, el nodo de consulta P11 está por encima de los nodos de la evidencia, así que aplicamos otra vez el nivel principal del procedimiento (mediante el teorema de Bayes):

$$p(\mathrm{P11}|\mathrm{P15}, \mathrm{P10}) = \frac{p(\mathrm{P15}, \mathrm{P10}|\mathrm{P11})p(\mathrm{P11})}{p(\mathrm{P15}, \mathrm{P10})} = k_1 p(\mathrm{P15}, \mathrm{P10}|\mathrm{P11})p(\mathrm{P11})$$

donde $k_1 = \frac{1}{p(\mathrm{P15}, \mathrm{P10})}$ y $p(\mathrm{P11})$ se da directamente en la red. El algoritmo finaliza con:

$$p(\mathrm{P15}, \mathrm{P10}|\mathrm{P11}) = p(\mathrm{P15}|\mathrm{P10}, \mathrm{P11})p(\mathrm{P10}|\mathrm{P11}) = p(\mathrm{P15}|\mathrm{P10}, \mathrm{P11})p(\mathrm{P10})$$

porque P10 y P11 son independientes.

Ahora se pueden agrupar todos los resultados, y los sumatorios, y k y k_1 se pueden calcular, para obtener la respuesta final de $p(\mathrm{Q}|\mathrm{P12}, \mathrm{P13}, \mathrm{P14}, \mathrm{P11})$.

La complejidad de los dos algoritmos, el de apoyo causal y el de apoyo evidencial, es lineal en relación al número de nodos en la red (sólo con poliárboles).

19.7.3. Apoyos causal y evidencial

Si disponemos de la evidencia por encima y por debajo de Q, como en:

$$p(\mathrm{Q}|\{\mathrm{P5}, \mathrm{P4}\}, \{\mathrm{P12}, \mathrm{P13}, \mathrm{P14}, \mathrm{P11}\})$$

separamos la evidencia en el subconjunto de apoyo causal ε^+, y en el de apoyo evidencial ε^-, y utilizamos el teorema de Bayes para escribir:

$$p(\mathrm{Q}|\varepsilon^+, \varepsilon^-) = \frac{p(\varepsilon^- | \mathrm{Q}, \varepsilon^+)p(\mathrm{Q} | \varepsilon^+)}{p(\varepsilon^- | \varepsilon^+)}$$

Como es habitual, tratamos $\frac{1}{p(\varepsilon^- | \varepsilon^+)} = k_2$ como un factor de normalización, y escribimos:

$$p(\mathrm{Q}|\varepsilon^+, \varepsilon^-) = k_2 p(\varepsilon^-|\mathrm{Q}, \varepsilon^+)p(\mathrm{Q}|\varepsilon^+)$$

Fijémonos en que Q separa-de ε^- de ε^+, de esta manera:

$$p(\mathrm{Q}|\varepsilon^+, \varepsilon^-) = k_2 p(\varepsilon^-|\mathrm{Q})p(\mathrm{Q}|\varepsilon^+)$$

Ya calculamos la primera probabilidad como parte del procedimiento descendente para calcular $p(\mathrm{Q}|\varepsilon^-)$; y la segunda probabilidad se calculó directamente con el procedimiento ascendente.

19.7.4. Un ejemplo numérico

Vamos a demostrar numéricamente la utilidad de estos métodos con un pequeño ejemplo, a partir del poliárbol abstracto que se muestra en la Figura 19.5. Lo que queremos es calcular $p(\mathrm{Q}|\mathrm{U})$.

Como es habitual en el razonamiento de diagnóstico, primero utilizamos el teorema de Bayes para obtener:

$$p(\mathrm{Q}|\mathrm{U}) = kp(\mathrm{U}|\mathrm{Q})p(\mathrm{Q}), \text{ donde } k = \frac{1}{p(\mathrm{U})}.$$

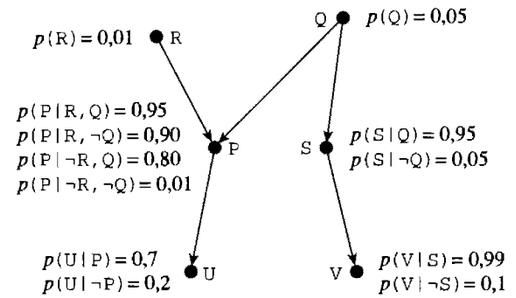


Figura 19.5.

Un pequeño poliárbol.

El algoritmo descendente calcula de forma recursiva:

$$p(\mathrm{U}|\mathrm{Q}) = \sum_{\mathrm{P}} p(\mathrm{U}|\mathrm{P})p(\mathrm{P}|\mathrm{Q})$$

$$p(\mathrm{P}|\mathrm{Q}) = \sum_{\mathrm{R}} p(\mathrm{P}|\mathrm{R}, \mathrm{Q})p(\mathrm{R})$$
$$= p(\mathrm{P}|\mathrm{R}, \mathrm{Q})p(\mathrm{R}) + p(\mathrm{P}|\neg\mathrm{R}, \mathrm{Q})p(\neg\mathrm{R})$$
$$= 0{,}95 \times 0{,}01 + 0{,}8 \times 0{,}99 = 0{,}80$$

$$p(\neg\mathrm{P}|\mathrm{Q}) = 0{,}20$$

$$p(\mathrm{U}|\mathrm{Q}) = p(\mathrm{U}|\mathrm{P}) \times 0{,}8 + p(\mathrm{U}|\neg\mathrm{P}) \times 0{,}2$$
$$= 0{,}7 \times 0{,}8 + 0{,}2 \times 0{,}2 = 0{,}60$$

$$p(\mathrm{Q}|\mathrm{U}) = k \times 0{,}6 \times 0{,}05 = k = 0{,}03$$

$$p(\neg\mathrm{Q}|\mathrm{U}) = kp(\mathrm{U}|\neg\mathrm{Q})p(\neg\mathrm{Q})$$

$$p(\mathrm{U}|\neg\mathrm{Q}) = \sum_{\mathrm{P}} p(\mathrm{U}|\mathrm{P})p(\mathrm{P}|\neg\mathrm{Q})$$

$$p(\mathrm{P}|\neg\mathrm{Q}) = \sum_{\mathrm{R}} p(\mathrm{P}|\mathrm{R}, \neg\mathrm{Q})p(\mathrm{R})$$
$$= p(\mathrm{P}|\mathrm{R}, \neg\mathrm{Q})p(\mathrm{R}) + p(\mathrm{P}|\neg\mathrm{R}, \neg\mathrm{Q})p(\neg\mathrm{R})$$
$$= 0{,}90 \times 0{,}01 + 0{,}01 \times 0{,}99 = 0{,}019$$

$$p(\neg\mathrm{P}|\neg\mathrm{Q}) = 0{,}98$$

$$p(\mathrm{U}|\neg\mathrm{Q}) = p(\mathrm{U}|\mathrm{P}) \times 0{,}019 + p(\mathrm{U}|\neg\mathrm{P}) \times 0{,}98$$
$$= 0{,}7 \times 0{,}019 + 0{,}2 \times 0{,}98 = 0{,}21$$

$$p(\neg\mathrm{Q}|\mathrm{U}) = k \times 0{,}21 \times 0{,}95 = k \times 0{,}20$$

De esta manera, $k = 4{,}35$, y, finalmente,

$$p(\mathrm{Q}|\mathrm{U}) = 4{,}35 \times 0{,}03 = 0{,}13$$

Los cálculos como los de este ejemplo se pueden organizar para evitar subcálculos redundantes. Uno de estos métodos se basa en utilizar lo que se denomina *eliminación del compartimiento* [Dechter, 1996].

Cuando la red no es un poliárbol, los procedimientos recursivos que acabamos de describir no finalizan, debido a los múltiples caminos (rutas) que hay entre los nodos. Se han propuesto otras técnicas para tratar estas redes más complejas. Entre ellas está un método de Monte Carlo (denominado *muestreo lógico* [Henrion, 1988]). En esta técnica, las probabilidades marginales de los nodos padres se utilizan para asignar valores aleatorios (como *Verdadero* o *Falso*) a dichos nodos. Luego, estos valores se utilizan junto con las TPC de sus descendientes para asignar valores aleatorios a dichos descendientes, y así se va recorriendo por la red. Finalmente, cada nodo en la red dispone de un valor. Este proceso se repite varias veces, y almacenamos todos los valores asignados a los nodos. En el límite de un número inifinito de intentos, los valores de los nodos serán consistentes con la probabilidad conjunta de los nodos especificados en la red, y con las TPC. Después de un número grande de intentos, podemos estimar $p(\mathrm{Q}, \mathrm{E})$, mediante la división del número de veces que Q y E han sido asignados a *Verdadero* entre el número de veces que E ha sido asignado a *Verdadero*. Está claro que se puede utilizar este mismo método para calcular la probabilidad conjunta de un conjunto de nodos de consulta dado un conjunto de nodos de la evidencia.

Otro método, denominado de *agrupamiento* [Lauritzen y Spiegelhalter, 1988], agrupa los nodos de la red en «meganodos», de tal modo que el grafo de los meganodos es un poliárbol. Los posibles valores de los meganodos son todas las combinaciones de los valores de sus nodos componentes. Entonces, se le puede aplicar el algoritmo del poliárbol, pero ahora tenemos varias TPC por cada meganodo, dadas las probabilidades condicionales de todos los valores de los meganodos, condicionadas por todos los valores de los nodos padre (los cuales pueden ser, a su vez, meganodos).

Lecturas y consideraciones adicionales

Hay varios libros de texto sobre probabilidades que se pueden consultar para suplir nuestro breve repaso; [Feller, 1968] es uno de ellos.

Algunos investigadores piensan que el razonamiento no monotónico se puede tratar mejor con los métodos probabilísticos, véase, por ejemplo, [Goldszmidt; Morris, y Pearl, 1990].

En IA, los trabajos sobre inferencia probabilística mediante redes bayesianas comenzaron con [Pearl, 1982a, y Kim y Pearl, 1983], quienes desarrollaron los algoritmos de «paso de mensajes» para los árboles y los poliárboles. El método que hemos descrito en este capítulo sobre poliárboles está basado en el de [Russell y Norvig, 1995, p. 447 y ss.]. Nuestro tratamiento sobre las redes bayesianas ha estado acotado a variables con valores discretos. También se ha desarrollado algún trabajo para variables aleatorias continuas; véase [Shachter y Kenley, 1989]. [Wellman 1990] ha investigado sobre redes «cualitativas».

Ya hemos citado el libro sobre inferencia probabilística de [Pearl, 1984]. [Neapolitan, 1990] es un libro de texto sobre el uso de los métodos probabilísticos en sistemas expertos. [Henrion, 1990] es un artículo introductorio sobre la inferencia probabilística en redes bayesianas. [Jensen, 1996] es un libro de texto sobre redes bayesianas, con el sistema HUGIN. [Neal, 1991], investiga las conexiones entre las redes bayesianas y las redes neuronales. David Heckerman, Michael Wellman, y Abe Mamdani fueron editores invitados en volumen especial sobre «Incertidumbre en IA» de la Communications of the ACM (vol. 38, n.º 3, marzo, 1995).

Las redes bayesianas se han utilizado en muchas aplicaciones de sistemas expertos. Un ejemplo típico es el PATHFINDER, un sistema que asiste a los patólogos en el diagnóstico de enfermedades del sistema linfático [Heckerman, 1991, y Heckerman y Nathwani, 1992]. Otro sistema es el CPCS-BN para medicina interna [Pradhan *et al.*, 1994], que tiene 448 nodos y 908 arcos, y que se ha visto evaluado favorablemente por los principales médicos del mundo en medicina interna.

Existen diversas alternativas a las redes bayesianas en el razonamiento con incertidumbre. El sistema experto MYCIN para diagnóstico médico y para recomendaciones de tratamiento utilizaba los *factores de certeza* [Shortliffe, 1976, y Buchanan y Shortliffe, 1984]. [Duda; Hart, y Nilsson, 1976] utilizaban índices de *suficiencia* y de *necesidad* en su sistema experto PROSPECTOR para la ayuda de exploración minera.

Otros métodos están basados en la lógica difusa y en la «teoría de la posibilidad» [Zadeh, 1975; Zadeh, 1978, y Elkan, 1993], o las funciones de creencia de Dempster-Shafer [Dempster, 1968, y Shafer, 1979]. [Nilsson, 1986], desarrolla una «lógica probabilística» y da citas sobre trabajos relacionados con la teoría de las probabilidades y las lógicas multivaluadas. Nuestro punto de vista en la actualidad es que, en muchas aplicaciones de los sistemas expertos, las redes bayesianas dominan a estos otros métodos; sin embargo, esta área aún sigue siendo un tema polémico.

Cuando el comportamiento humano se enfrenta a la incertidumbre, puede ser bastante inconsistente [Tversky y Kahneman, 1982], y puede generar modelos poco útiles para la ingeniería.

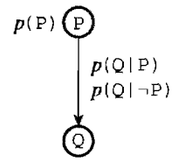
[Shafer y Pearl, 1990], es una colección de artículos sobre el razonamiento con incertidumbre. Las actas de las conferencias anuales de *Uncertainty in Artificial Intelligence* (UAI) contienen las descripciones de las investigaciones actuales. La *International Journal of Approximate Reasoning*, así como otras revistas y conferencias sobre IA, publican artículos importantes.

19.1. Supongamos que unas bolas coloreadas se colocan en tres cajas, *C*1, *C*2 y *C*3, del siguiente modo:

	*C*1	*C*2	*C*3
Roja	2	4	3
Blanca	3	2	4
Azul	6	3	3

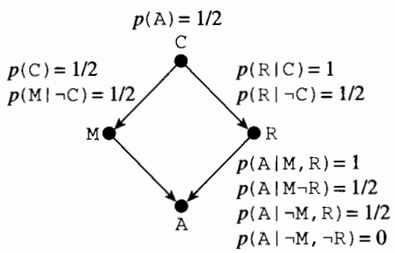
Se selecciona aleatoriamente una bola de cualquiera de las tres cajas. La bola es roja. ¿Cuáles son las probabilidades de que la caja seleccionada sea la *C*1, la *C*2 o la *C*3? Razone la respuesta.

19.2. Considere la red de creencia que se muestra.



1. Derive una expresión para la probabilidad de $\mathrm{P} \supset \mathrm{Q}$.
2. ¿Cuándo son iguales $p(\mathrm{P} \supset \mathrm{Q})$ y $p(\mathrm{Q}|\mathrm{P})$?
3. Asuma que no se conoce la tabla de probabilidades condicionales de la red. En su lugar, todo lo que se conoce son los valores de $p(\mathrm{P})$ y de $p(\mathrm{P} \supset \mathrm{Q})$. ¿Qué se puede decir acerca del valor de $p(\mathrm{Q})$?

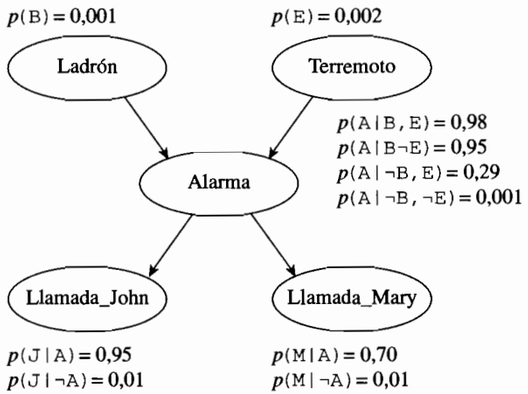
19.3. La comisión de admisión de un instituto está intentando determinar la probabilidad de que un candidato admitido realmente esté cualificado. Las probabilidades relevantes se dan en la red que se muestra abajo. Calcule $p(C|A)$.



C = el candidato está cualificado
M = el candidato tiene una buena media académica
R = el candidato tiene excelentes recomendaciones
A = el candidato es admitido

$p(C|A) = ?$

19.4. (Cortesía de Judea Pearl, un residente en una zona sísmica.) La red de creencia que se muestra formaliza la siguiente situación. Se tiene un nuevo sistema de alarma antirrobo instalado en casa. Este sistema es bastante fiable en detectar un allanamiento de morada, pero también se activa a veces al más pequeño terremoto. También se tienen dos vecinos, John y Mary, que han prometido llamar al trabajo si oyen la alarma. Las llamadas de John cuando oye la alarma son absolutamente fiables, pero a veces confunde la alarma con el teléfono, e igualmente llama. Por otro lado, a Mary le gusta la música ruidosa, y algunas veces se olvida de la alarma.



Para ejercitar la habilidad de trabajar con probabilidades conjuntas definidas mediante redes de creencia, calcule la probabilidad conjunta de que no llame ni John ni Mary, y que haya un terremoto y un allanamiento de morada, es decir, calcule $p(\neg J, \neg M, B, E)$.

19.5. En una galaxia lejana, muy lejana, el 90 por 100 de los taxis son verdes, y el 10 por 100 azules. Sucede un accidente en el que participa un taxi; se presume que el porcentaje de accidentes de taxis verdes es el mismo que el de azules. Un juzgado analiza el accidente, y un periodista de un periódico que estuvo en la escena del accidente dice, «El taxi era azul». Este periodista es, por lo general, fiable, de hecho, sus enunciados son correctos en un 80 por 100 de las veces. De todo esto se saca, que el taxi que participó en el accidente era azul (o verde), y la probabilidad de que nuestro testigo dijera «azul» (o «verde») es 0,8. ¿Cuál es la probabilidad de que el taxi fuera azul, dada la afirmación del periodista?

19.6. A Orville, el robot malabarista, se le caen las bolas bastante a menudo cuando su batería está baja. En tests anteriores, se ha determinado que la probabilidad de que se le caiga una bola cuando la batería está baja es 0,9. Siempre que la batería no está baja, la probabilidad de que se le caiga una bola sólo es 0,01. La batería se cargó no hace mucho, y nuestra mejor conjetura (antes de ver el último registro malabar de Orville) es que las probabilidades de que la batería esté baja es de 10 contra 1. Un observador, con algún tipo de sistema de visión poco fiable, informa de que a Orville se le ha caido una bola. La fiabilidad del observador se da por las siguientes probabilidades:

p(el observador dice que a Orville se le ha caido una bola | se le ha caido la bola a Orville) = 0,9
p(el observador dice que a Orville se le ha caido una bola | no se le ha caido la bola a Orville) = 0,2

Dibuje las red bayesiana y calcule la probabilidad de que la batería esté baja dado el informe del observador.