

Notación de probabilidad

Se usarán mayúsculas para indicar variables estocásticas y minúsculas para indicar los valores que pueden tomar.

$$P(A = \text{verdadero}) = P(A = a) = P(a)$$

$$P(A = \text{falso}) = P(A = \neg a) = P(\neg a)$$

$$P(a \wedge b \wedge c) \equiv P(a, b, c) \equiv P(a \, b \, c)$$

$$P(\neg a) \equiv P(a') \equiv P(\bar{a})$$

Notación vectorial

Usaremos la notación vectorial de la siguiente manera.

$$\mathbf{P}(A) = \langle P(a), P(\neg a) \rangle$$

Por ejemplo, $\mathbf{P}(X, Y) = \mathbf{P}(X|Y)P(Y)$ es equivalente a

$$P(\neg x, \neg y) = P(\neg x|\neg y)P(\neg y)$$

$$P(\neg x, y) = P(\neg x|y)P(y)$$

$$P(x, \neg y) = P(x|\neg y)P(\neg y)$$

$$P(x, y) = P(x|y)P(y)$$

Normalización

Para obtener $\mathbf{P}(X|y)$ se puede hacer lo siguiente:

$$P(x|y) = \frac{P(x, y)}{P(y)}$$

$$P(\neg x|y) = \frac{P(\neg x, y)}{P(y)}$$

Otra forma es la siguiente

$$P(x|y) = \alpha P(x, y)$$

$$P(\neg x|y) = \alpha P(\neg x, y)$$

donde α representa la constante de *normalización* que sea necesaria para lograr que $P(x|y) + P(\neg x|y) = 1$.

Probabilidad condicional

Probabilidad condicional, o probabilidad posterior, es la probabilidad de a dado que se sabe b . Se define como

$$P(a|b) \equiv \frac{P(a \wedge b)}{P(b)}$$

De lo anterior se puede obtener lo siguiente:

$$P(a \wedge b) = P(a|b) P(b) = P(b|a) P(a)$$

Los axiomas de probabilidad

1. $0 \leq P(a) \leq 1$
2. $P(\text{verdadero}) = 1$ y $P(\text{falso}) = 0$
3. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$

Probabilidad conjunta y probabilidad marginal

La distribución de probabilidad conjunta es una tabulación de la probabilidad de todos los valores que pueden tomar las variables aleatorias y la probabilidad para cada combinación.

La distribución de probabilidad marginal para un subconjunto de variables se obtiene sumando sobre las demás variables.

Ejemplo

Distribución de probabilidad conjunta:

	toothache		\neg toothache	
	catch	\neg catch	catch	\neg catch
cavity	0.108	0.012	0.072	0.008
\neg cavity	0.016	0.064	0.144	0.576

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064$$

$$P(\neg \text{toothache}) = 0.072 + 0.008 + 0.144 + 0.576$$

$$P(\text{catch} \wedge \neg \text{cavity}) = 0.016 + 0.144$$

Teorema de Bayes

$$P(b|a) = \frac{P(a|b) P(b)}{P(a)}$$

Recuérdese que

$$P(b|a) = \frac{P(a \wedge b)}{P(a)}$$

Ejemplo

Supongamos que se sabe que lo siguiente:

- Meningitis causa dolor de cuello en 50 % de los casos, es decir, $P(s|m) = 0.5$
- La probabilidad de tener meningitis es $P(m) = 1/50,000$
- La probabilidad de tener dolor de cuello es $P(s) = 1/20$

$$P(m|s) = \frac{P(s|m) P(m)}{P(s)} = 0.0002$$

Independencia

a y b son independientes, si y sólo si se cumplen las siguientes condiciones equivalentes

$$P(a|b) = P(a)$$

$$P(b|a) = P(b)$$

$$P(a \wedge b) = P(a) P(b)$$

Definiciones de Pearl

- Prior odds

$$O(h) = \frac{P(h)}{P(\neg h)} = \frac{P(h)}{1 - P(h)}$$

- Likelihood ratio

$$L(e|h) = \frac{P(e|h)}{P(e|\neg h)}$$

- Posterior odds

$$O(h|e) = \frac{P(h|e)}{P(\neg h|e)} = L(e|h) o(h)$$

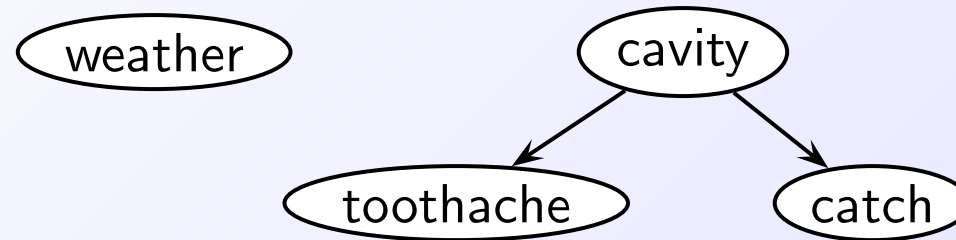
donde h es una hipótesis y e es una evidencia.

Redes bayesianas

Una *red bayesiana* es una representación gráfica que representa las dependencias entre variables y da una especificación consisa de cualquier distribución de probabilidad conjunta completa.

Las relaciones de causa/efecto (o hipótesis/evidencia) se representan con arcos dirigidos de la causa al efecto.

Ejemplo



En este ejemplo, cavity es evidencia de toothache y de catch. Además, weather es independiente de las demás variables.

Inferencia en redes Bayesianas

La tarea básica de un sistema de inferencia probabilística es calcular las distribuciones de probabilidad posteriores para un conjunto de variables dado un evento observado, es decir, dada una asignación de valores a un conjunto de variables de evidencia.

Inferencia exacta por enumeración

La probabilidad de la hipótesis H dada la evidencia e :

$$P(H|e) = \frac{P(H, e)}{P(e)}$$

En caso de varias evidencias e_1, e_2, \dots, e_n :

$$P(H|e_1, e_2, \dots, e_n) = \frac{P(e_1|H)P(e_2|H) \cdots P(e_n|H)}{P(e_1)P(e_2) \cdots P(e_n)} P(H)$$

O-ruidoso (Noisy-OR)

El O-ruidoso permite que haya incertidumbre acerca de la capacidad de cada padre para hacer que el hijo sea verdadero—la relación causal entre padre e hijo puede ser *inhibida*, de manera que un paciente puede tener gripe, pero no tener fiebre.

Dos suposiciones:

1. Todas las causas posibles han sido listada.
2. La inhibición de cada padre es independiente de las inhibiciones de los demás padres.

Ejemplo de O-ruidoso

Se tienen las siguientes probabilidades:

$$P(\neg \text{fiebre} | \text{gripe}, \neg \text{influenza}, \neg \text{malaria}) = 0.6$$

$$P(\neg \text{fiebre} | \neg \text{gripe}, \text{influenza}, \neg \text{malaria}) = 0.2$$

$$P(\neg \text{fiebre} | \neg \text{gripe}, \neg \text{influenza}, \text{malaria}) = 0.1$$

Ejemplo de O-ruidoso

Se tienen las siguientes probabilidades:

$$P(\neg \text{fiebre} | \text{gripe}, \neg \text{influenza}, \neg \text{malaria}) = 0.6$$

$$P(\neg \text{fiebre} | \neg \text{gripe}, \text{influenza}, \neg \text{malaria}) = 0.2$$

$$P(\neg \text{fiebre} | \neg \text{gripe}, \neg \text{influenza}, \text{malaria}) = 0.1$$

gripe	influenza	malaria	$P(\text{fiebre})$	$P(\neg \text{fiebre})$
<i>F</i>	<i>F</i>	<i>F</i>	0.0	1.0
<i>F</i>	<i>F</i>	<i>T</i>	0.9	0.1
<i>F</i>	<i>T</i>	<i>F</i>	0.8	0.2
<i>F</i>	<i>T</i>	<i>T</i>	0.98	$0.02 = 0.2 \times 0.1$
<i>T</i>	<i>F</i>	<i>F</i>	0.4	0.6
<i>T</i>	<i>F</i>	<i>T</i>	0.9	$0.06 = 0.6 \times 0.1$
<i>T</i>	<i>T</i>	<i>F</i>	0.8	$0.12 = 0.6 \times 0.2$
<i>T</i>	<i>T</i>	<i>T</i>	0.98	$0.012 = 0.6 \times 0.2 \times 0.1$

Ejemplo mínimo

A

$P(\neg a)$	$P(a)$
0.5	0.5



A	$P(\neg b)$	$P(b)$
F	1	0
T	0.1	0.9

Función de distribución de probabilidad conjunta

$$P(\neg a, \neg b) = P(\neg b|\neg a)P(\neg a) = 1 \times 0.5 = 0.5$$

$$P(\neg a, b) = P(b|\neg a)P(\neg a) = 0 \times 0.5 = 0$$

$$P(a, \neg b) = P(\neg b|a)P(a) = 0.1 \times 0.5 = 0.05$$

$$P(a, b) = P(b|a)P(a) = 0.9 \times 0.5 = 0.45$$

	$P(\neg b)$	$P(b)$
$P(\neg a)$	0.5	0
$P(a)$	0.05	0.45

Probabilidades anteriores

$$P(\neg a) = P(\neg a, \neg b) + P(\neg a, b) = 0.5$$

$$P(a) = P(a, \neg b) + P(a, b) = 0.5$$

$$P(\neg b) = P(\neg a, \neg b) + P(a, \neg b) = 0.55$$

$$P(b) = P(\neg a, b) + P(a, b) = 0.45$$

	$P(\neg b)$	$P(b)$	
$P(\neg a)$	0.5	0	0.5
$P(a)$	0.05	0.45	0.5
	0.55	0.45	

Probabilidades posteriores

$$\begin{aligned}P(\neg a|\neg b) &= \frac{P(\neg a, \neg b)}{P(\neg b)} = \frac{0.5}{0.55} \\&= \frac{P(\neg b|\neg a)P(\neg a)}{P(\neg b)} = \frac{1 \times 0.5}{0.55} = 0.9091\end{aligned}$$

$$\begin{aligned}P(a|\neg b) &= \frac{P(a, \neg b)}{P(\neg b)} = \frac{0.05}{0.55} \\&= \frac{P(\neg b|a)P(a)}{P(\neg b)} = \frac{0.1 \times 0.5}{0.55} = 0.0909\end{aligned}$$

$$P(\neg a|b) = \frac{P(\neg a, b)}{P(b)} = \frac{0}{0.45} = \frac{P(b|\neg a)P(\neg a)}{P(\neg b)} = \frac{0 \times 0.5}{0.55} = 0$$

$$P(a|b) = \frac{P(a, b)}{P(b)} = \frac{0.45}{0.45} = \frac{P(b|a)P(a)}{P(b)} = \frac{0.9 \times 0.5}{0.45} = 1$$

Bayesian Networks Toolbox

El BNT (Bayesian Networks Toolbox) permite implementar redes bayesianas en MATLAB. Los siguientes comandos implementan el ejemplo mínimo anterior.

Primero, definimos la topología de la red, y se cargan las distribuciones de probabilidad condicionales:

```
N=2;  
dag=zeros(N,N);  
dag(1,2)=1;  
node_sizes=2*ones(1,N);  
bnet=mk_bnet(dag,node_sizes);  
bnet.CPD{1}=tabular_CPD(bnet,1,[0.5 0.5]);  
bnet.CPD{2}=tabular_CPD(bnet,2,[1 0.1 0 0.9]);
```

Los nodos deben numerarse a partir de 1, y de nodos padres a nodos hijos.

Inferencias en BNT

Para realizar inferencias en BNT es necesario escoger una máquina de inferencias. La evidencia se carga en una lista (en este caso están vacíos los elementos).

```
engine = jtree_inf_engine(bnet);  
evidence = cell(1,N);  
[engine, loglik] = enter_evidence(engine, evidence);
```

Las probabilidades marginales se obtienen con el comando `marginal_nodes`. En este caso obtenemos la probabilidad conjunta:

```
marg = marginal_nodes(engine, [1 2]);  
marg.T
```


Probabilidades anteriores y posteriores

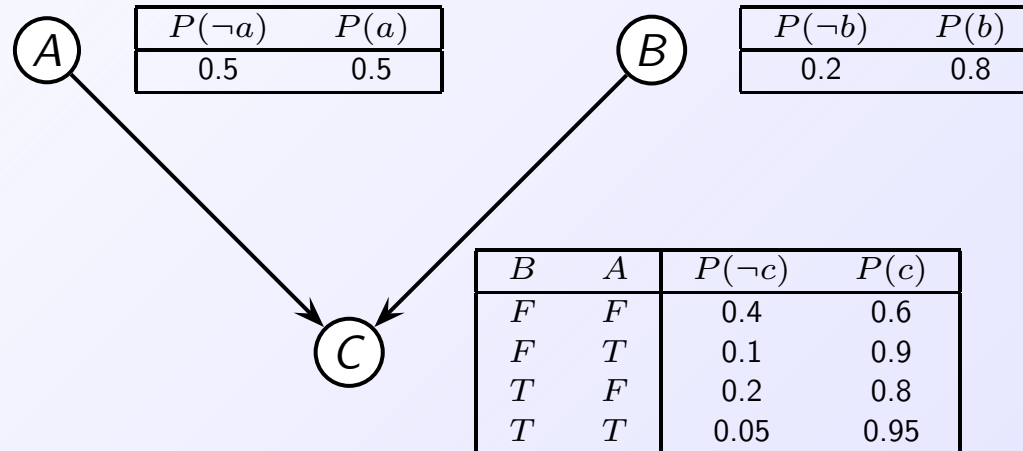
```
marg = marginal_nodes(engine, 1);  
marg.T
```

```
marg = marginal_nodes(engine, 2);  
marg.T
```

```
evidence=cell(1,N);  
evidence{2}=1;  
[engine, loglik] = enter_evidence(engine, evidence);  
marg = marginal_nodes(engine,1);  
marg.T
```

```
evidence{2}=2;  
[engine, loglik] = enter_evidence(engine, evidence);  
marg = marginal_nodes(engine,1);  
marg.T
```

Otro ejemplo de red bayesiana



Función de distribución de probabilidad conjunta

	$P(\neg b)$	$P(b)$
$P(\neg a)$	0.1	0.4
$P(a)$	0.1	0.4

$$P(\neg a, \neg b, \neg c) = P(\neg c|\neg a, \neg b)P(\neg a, \neg b) = 0.4 \times 0.1 = 0.04$$

$$P(\neg a, \neg b, c) = P(c|\neg a, \neg b)P(\neg a, \neg b) = 0.6 \times 0.1 = 0.06$$

$$P(a, \neg b, \neg c) = P(\neg c|a, \neg b)P(a, \neg b) = 0.1 \times 0.1 = 0.01$$

$$P(a, \neg b, c) = P(c|a, \neg b)P(a, \neg b) = 0.9 \times 0.1 = 0.09$$

$$P(\neg a, b, \neg c) = P(\neg c|\neg a, b)P(\neg a, b) = 0.2 \times 0.4 = 0.08$$

$$P(\neg a, b, c) = P(c|\neg a, b)P(\neg a, b) = 0.8 \times 0.4 = 0.32$$

$$P(a, b, \neg c) = P(\neg c|a, b)P(a, b) = 0.05 \times 0.4 = 0.02$$

$$P(a, b, c) = P(c|a, b)P(a, b) = 0.95 \times 0.4 = 0.38$$

Probabilidades anteriores

$$\begin{aligned}P(a) &= P(a, \neg b, \neg c) + P(a, \neg b, c) + P(a, b, \neg c) + P(a, b, c) \\&= 0.01 + 0.09 + 0.02 + 0.38 = 0.5\end{aligned}$$

$$\begin{aligned}P(b) &= P(\neg a, b, \neg c) + P(\neg a, b, c) + P(a, b, \neg c) + P(a, b, c) \\&= 0.08 + 0.32 + 0.02 + 0.38 = 0.8\end{aligned}$$

$$\begin{aligned}P(c) &= P(\neg a, \neg b, c) + P(\neg a, b, c) + P(a, \neg b, c) + P(a, b, c) \\&= 0.06 + 0.09 + 0.32 + 0.38 = 0.85\end{aligned}$$

$$\begin{aligned}P(a, b) &= P(a, b, \neg c) + P(a, b, c) \\&= 0.02 + 0.38 = 0.4\end{aligned}$$

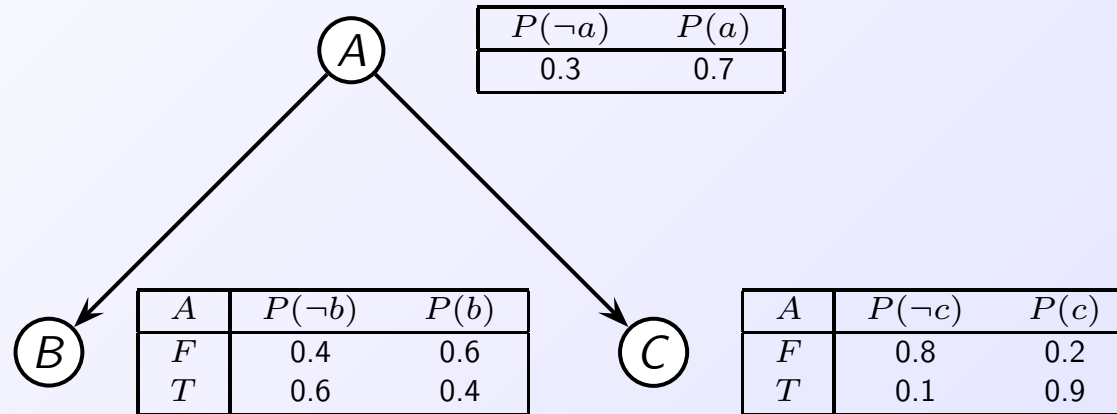
$$\begin{aligned}P(a, c) &= P(a, \neg b, c) + P(a, b, c) \\&= 0.09 + 0.38 = 0.47\end{aligned}$$

$$\begin{aligned}P(b, c) &= P(\neg a, b, c) + P(a, b, c) \\&= 0.32 + 0.38 = 0.7\end{aligned}$$

Probabilidades posteriores

$$\begin{aligned}P(a, b|c) &= \frac{P(a, b, c)}{P(c)} = \frac{0.38}{0.85} = 0.4471 \\&= \frac{P(c|a, b)P(a, b)}{P(c)} = \frac{0.95 \times 0.4}{0.85} = 0.4471 \\P(a|b, c) &= \frac{P(a, b, c)}{P(b, c)} = \frac{0.38}{0.7} = 0.5429\end{aligned}$$

Un ejemplo más

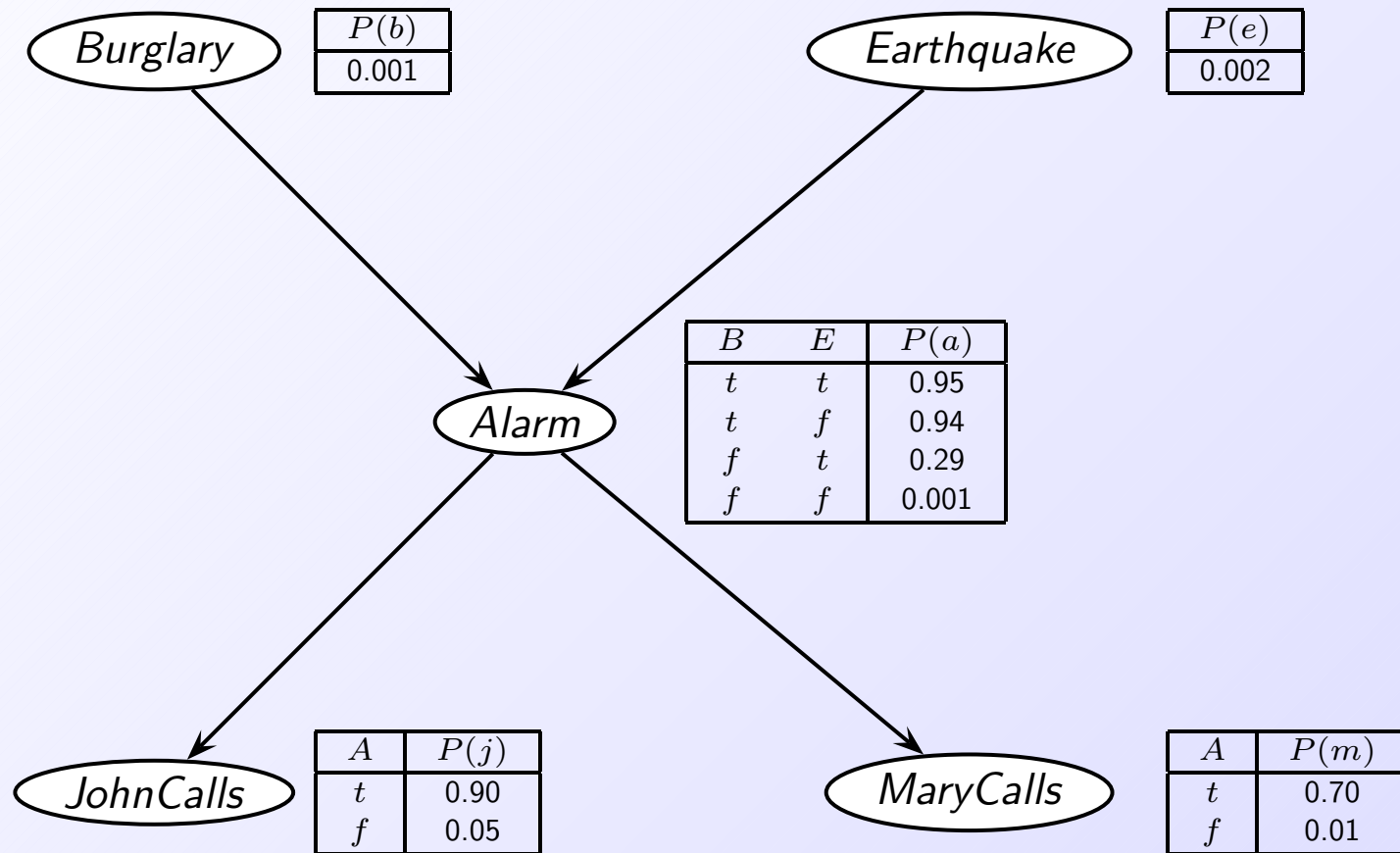


Probabilidad posterior

$$P(a|c) = \frac{P(c|a)P(a)}{P(c)} = \frac{0.9 \times 0.7}{0.69} = 0.9130$$

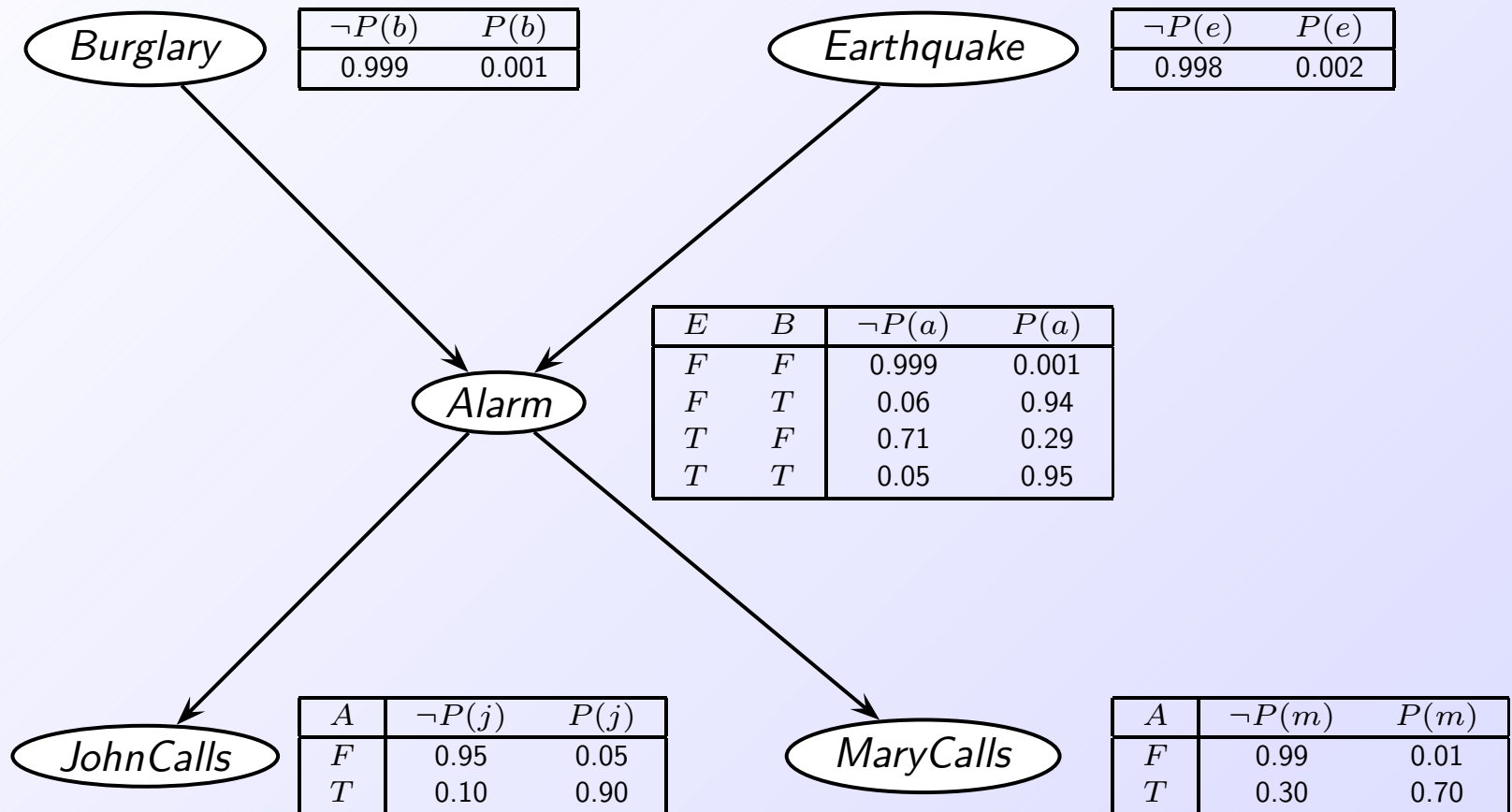
Ejemplo de Pearl

Como aparece en el texto:



Ejemplo de Pearl

Como se debe cargar en BNT:



BNT (Bayesian Networks Toolbox)

Creación de la red

```
N = 5;  
dag = zeros(N,N);  
B = 1; E = 2; A = 3; J = 4; M = 5;  
dag(B,A) = 1;  
dag(E,A) = 1;  
dag(A,[J M]) = 1;  
node_sizes = 2*ones(1,N);  
bnet = mk_bnet(dag, node_sizes);
```

Se cargan las distribuciones de probabilidad conjuntas:

```
bnet.CPD{B} = tabular_CPD(bnet, B, [0.001]);  
bnet.CPD{E} = tabular_CPD(bnet, E, [0.002]);  
bnet.CPD{A} = tabular_CPD(bnet, A, ...  
    [0.001 0.29 0.94 0.95]);  
bnet.CPD{J} = tabular_CPD(bnet, J, [0.05 0.90]);  
bnet.CPD{M} = tabular_CPD(bnet, M, [0.01 0.70]);
```

Se escoge una máquina de inferencia:

```
engine = jtree_inf_engine(bnet);
```

Probabilidad de robo dado que sonó la alarma

Se carga la evidencia ($A = 2$):

```
evidence = cell(1,N);  
evidence{A} = 2;  
[engine, loglik] = enter_evidence(engine, evidence);
```

Se calcula la probabilidad marginal de robo:

```
m = marginal_nodes(engine, B);  
m.T
```

que produce el resultado:

```
ans =
```

```
0.9080
```

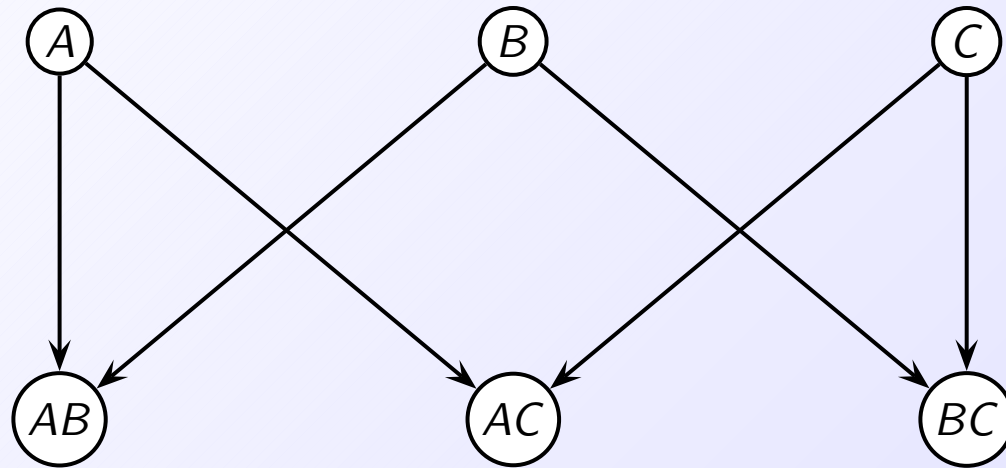
```
0.0920
```

es decir, $P(B|A) = 0.9080$.

Probabilidad de robo dado que John llamó

```
evidence = cell(1,N);  
evidence{J} = 2;  
[engine, loglik] = enter_evidence(engine, evidence);  
m = marginal_nodes(engine, B);  
m.T
```

Problema 14.12 del texto



- A , B , y C representan las calidades de los equipos, y pueden tomar valores en $\{0, 1, 2, 3\}$.
- AB , AC , y BC representan los resultados de los juegos entre los equipos, y pueden tomar valores en $\{p, e, g\}$ donde p representa que el primer equipo pierda, e representa empate, y g representa que el primer equipo gane.

Probabilidades anteriores

Asumimos que todos los equipos son iguales inicialmente, es decir,

$P(A = 0)$	$P(A = 1)$	$P(A = 2)$	$P(A = 3)$
0.25	0.25	0.25	0.25

Y de la misma manera para B y C .

Probabilidades condicionales

Escojemos la probabilidad de un resultado de un juego en términos las diferencias de calidades.

diferencia	$P(p)$	$P(e)$	$P(g)$
-3	0.55	0.40	0.05
-2	0.40	0.50	0.10
-1	0.30	0.55	0.15
0	0.20	0.60	0.20
1	0.15	0.55	0.30
2	0.10	0.50	0.40
3	0.05	0.40	0.55

De aquí podemos obtener las probabilidades condicionales $\mathbf{P}(AB|A, B)$, $\mathbf{P}(AC|A, C)$, y $\mathbf{P}(BC|B, C)$.