人工智能：不确定知识表示和推理 II

饶洋辉

计算机学院,

中山大学

raoyangh@mail.sysu.edu.cn

http://cse.sysu.edu.cn/node/2471

课件来源：中山大学刘咏梅教授；多伦多大学Sheila McIlraith教授；浙江大学吴飞教授；海军工程大学贲可荣教授等

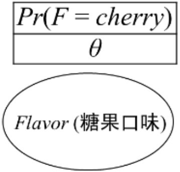
目录

贝叶斯网络的参数学习

- 条件概率表的学习方法:
  - 极大似然
  - 期望-最大化

贝叶斯网络的参数学习

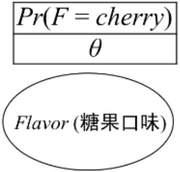
- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化
- 示例1：
  - 某超市销售的糖果有2种口味（*cherry*和*lime*）
  - 假设尝到*c*颗*cherry*口味的和*l*颗*lime*口味的



S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化



- 示例1：
  - 某超市销售的糖果有2种口味（*cherry* 和*lime*）
  - 假设尝到*c*颗*cherry*口味的和*l*颗*lime*口味的

记上述观察到的数据为*d*

$$Pr(d \mid h_\theta) = \theta^c (1-\theta)^l$$

$$\log Pr(d \mid h_\theta) = c \log\theta + l \log(1-\theta)$$

$$d\left(\log Pr(d \mid h_\theta)\right)/d\theta = c/\theta - l/(1-\theta)$$

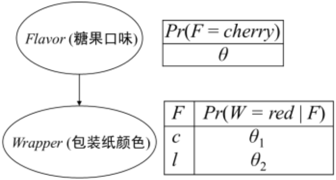
$$c/\theta - l/(1-\theta) = 0 \;\Rightarrow\; \theta = c/(c+l)$$

S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化

- 示例2：
  - 某超市销售的糖果有2种口味（*cherry*和*lime*），包装纸的颜色分为绿色和红色
  - 糖果的口味决定了包装纸的颜色
  - 假设尝到c颗*cherry*口味的（其中g_c颗为绿色包装纸，r_c颗为红色包装纸），以及l颗*lime*口味的（其中g_l颗为绿色包装纸，r_l颗为红色包装纸）



S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化

- 示例2：
  - 某超市销售的糖果有2种口味（*cherry*和*lime*），包装纸的颜色分为绿色和红色
  - 糖果的口味决定了包装纸的颜色
  - 假设尝到c颗*cherry*口味的（其中g_c颗为绿色包装纸，r_c颗为红色包装纸），以及l颗*lime*口味的（其中g_l颗为绿色包装纸，r_l颗为红色包装纸）

$Pr(F = cherry)$
θ

F	$Pr(W = red \mid F)$
c	θ_1
l	θ_2

$$Pr(d \mid h_{\theta,\theta_1,\theta_2}) = \theta^c \theta_1^{r_c} (1-\theta_1)^{g_c} (1-\theta)^l \theta_2^{r_l} (1-\theta_2)^{g_l}$$

$$c/\theta - l/(1-\theta) = 0 \;\Rightarrow\; \theta = c/(c+l)$$

$$r_c/\theta_1 - g_c/(1-\theta_1) = 0 \;\Rightarrow\; \theta_1 = r_c/(r_c+g_c)$$

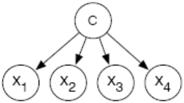
$$r_l/\theta_2 - g_l/(1-\theta_2) = 0 \;\Rightarrow\; \theta_2 = r_l/(r_l+g_l)$$

S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化
- 示例3：
  - 不可观测的k值随机变量C
  - 可观测的二值属性X_1至X_4

Model



Data

X_1	X_2	X_3	X_4
t	f	t	t
f	t	t	f
f	f	t	t
	$\dots$		

Probabilities

$P(C)$
$P(X_1|C)$
$P(X_2|C)$
$P(X_3|C)$
$P(X_4|C)$

S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化

- 示例3：
  - 不可观测的k值随机变量C
  - 可观测的二值属性X_1至X_4

假设k = 3。期望步：随机初始化条件概率表，通过贝叶斯定理，可以计算每行数据的后验概率，例如$Pr(C = 1 \mid X_1 = t, X_2 = f, X_3 = t, X_4 = t)$、$Pr(C = 2 \mid X_1 = t, X_2 = f, X_3 = t, X_4 = t)$、$Pr(C = 3 \mid X_1 = t, X_2 = f, X_3 = t, X_4 = t)$，假设分别为0.407、0.121、0.472，由此可以得到完备化的数据实例。

X_1	X_2	X_3	X_4	Count
⋮	⋮	⋮	⋮	⋮
t	f	t	t	100
⋮	⋮	⋮	⋮	⋮

$\longrightarrow$

$A[X_1, \ldots, X_4, C]$

X_1	X_2	X_3	X_4	C	Count
⋮	⋮	⋮	⋮	⋮	⋮
t	f	t	t	1	40.7
t	f	t	t	2	12.1
t	f	t	t	3	47.2
⋮	⋮	⋮	⋮	⋮	⋮

S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

贝叶斯网络的参数学习

- 条件概率表的学习方法：
  - 极大似然
  - 期望-最大化

- 示例3：
  - 不可观测的k值随机变量C
  - 可观测的二值属性X_1至X_4

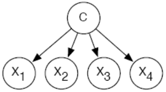
假设$k = 3$。最大化步：通过期望步得到的完备化数据实例，采用极大似然法，重新计算条件概率表。

$$P(C=v_i) = \frac{\sum_{t \models C=v_i} Count(t)}{\sum_t Count(t)}$$

$$P(X_k = v_j | C=v_i) = \frac{\sum_{t \models C=v_i \wedge X_k=v_j} Count(t)}{\sum_{t \models C=v_i} Count(t)}$$

X_1	X_2	X_3	X_4	C	$Count$
⋮	⋮	⋮	⋮	⋮	⋮
t	f	t	t	1	40.7
t	f	t	t	2	12.1
t	f	t	t	3	47.2
⋮	⋮	⋮	⋮	⋮	⋮

$\longrightarrow$



S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

目录

- 背景
- 贝叶斯网络的表示
- 贝叶斯网络的构建
- 贝叶斯网络的参数学习
- 贝叶斯网络中的D-分离
- 贝叶斯网络的推理

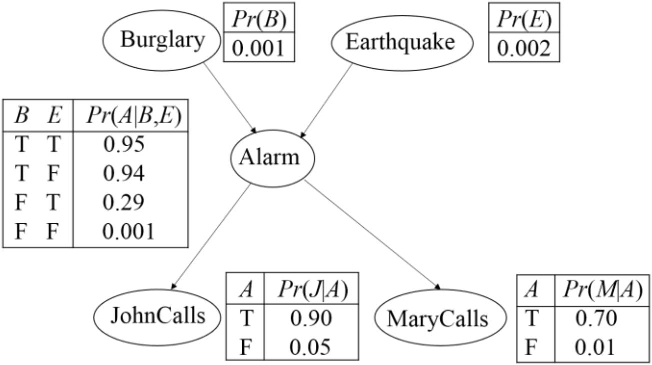
贝叶斯网络中的D-分离

- 对于一个有向无环图，D-分离（D-Separation）是一种用来判断其变量是否条件独立的图形化方法。
- 在基于贝叶斯网络的不确定性知识推理中，采用D-分离方法可以简化概率计算，提高运行速度。
- 示例：
  - 小偷（Burglar）会引发警报（Alarm）
  - 地震（Earthquake）会引发警报（Alarm）
  - 警报（Alarm）会引起邻居约翰打电话（JohnCalls）
  - 警报（Alarm）会引起邻居玛丽打电话（MaryCalls）

S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition, Prentice Hall, 2009.

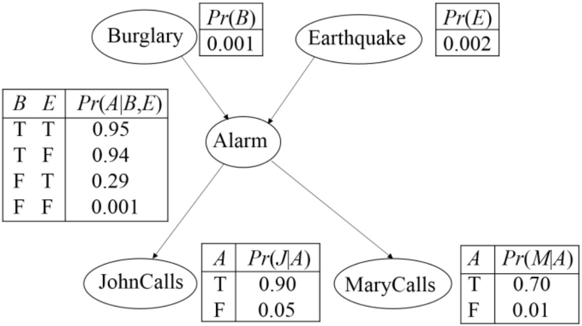
贝叶斯网络中的D-分离

- 示例：



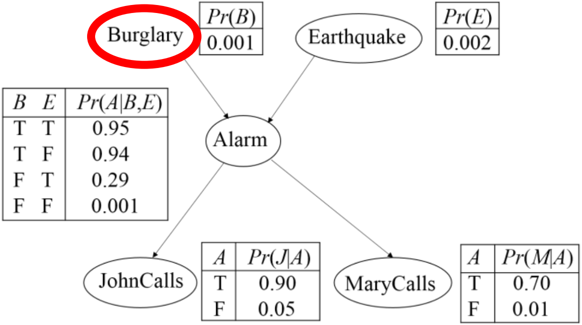
贝叶斯网络中的D-分离

- 间接的因果作用：



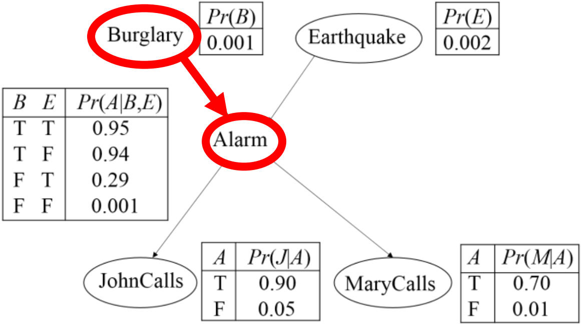
贝叶斯网络中的D-分离

- 间接的因果作用：



贝叶斯网络中的D-分离

- 间接的因果作用：



$Pr(B)$
0.001

$Pr(E)$
0.002

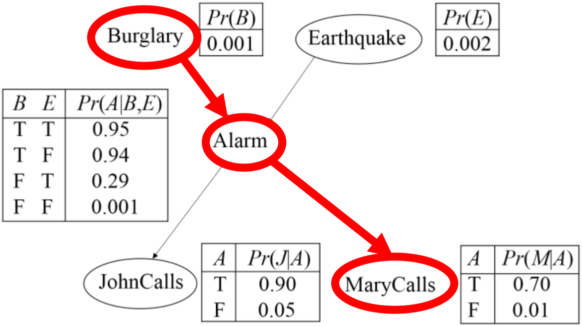
B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

贝叶斯网络中的D-分离

- 间接的因果作用：



$Pr(B)$
0.001

$Pr(E)$
0.002

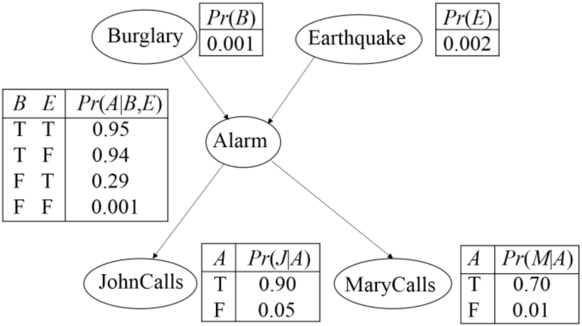
B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

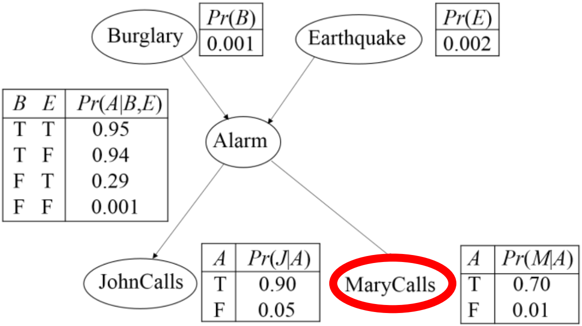
贝叶斯网络中的D-分离

- 间接的证据作用:



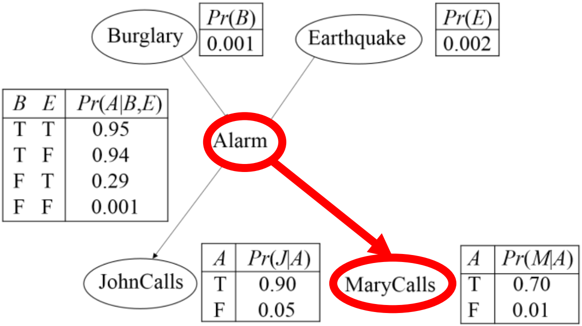
贝叶斯网络中的D-分离

- 间接的证据作用:



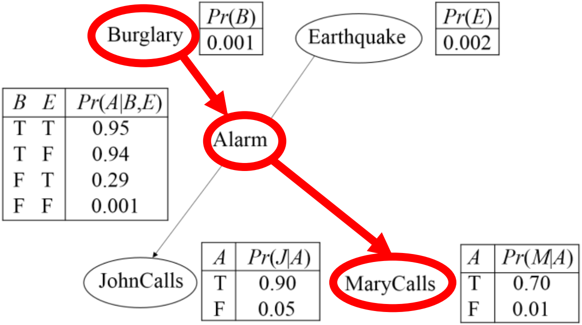
贝叶斯网络中的D-分离

- 间接的证据作用：



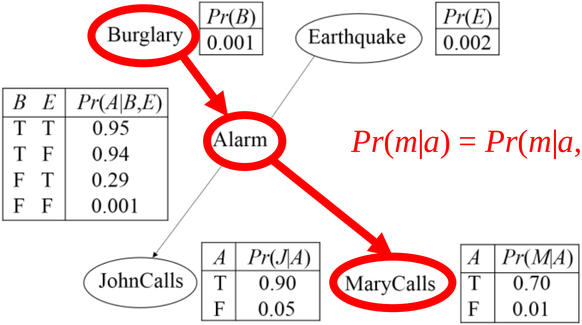
贝叶斯网络中的D-分离

- 间接的证据作用：



贝叶斯网络中的D-分离

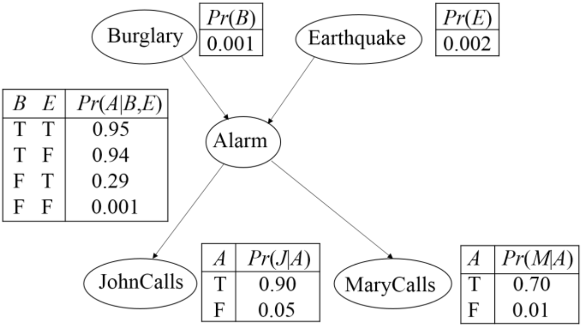
- 间接的因果/证据作用：给定A的前提下，M与B条件独立。



$Pr(m|a) = Pr(m|a,b) = 0.7$

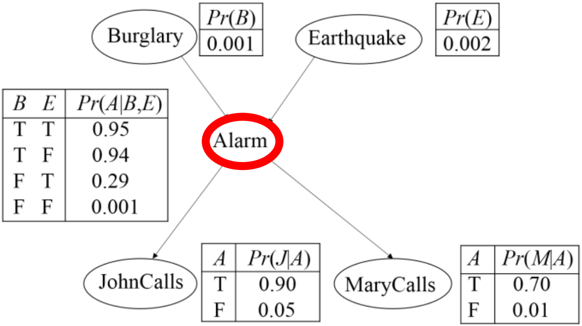
贝叶斯网络中的D-分离

- 共同的原因:



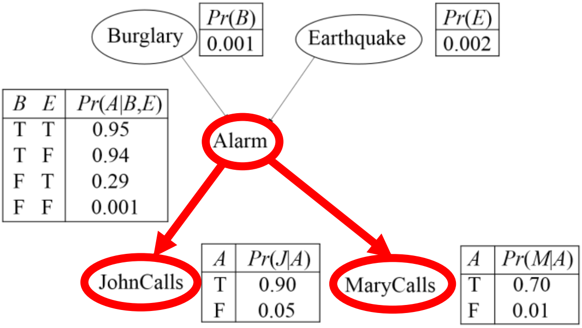
贝叶斯网络中的D-分离

- 共同的原因：



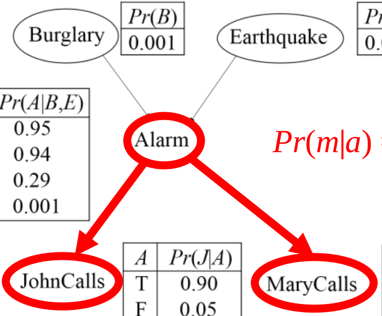
贝叶斯网络中的D-分离

- 共同的原因：



贝叶斯网络中的D-分离

- 共同的原因：给定A的前提下，J与M条件独立。



$Pr(B)$
0.001

$Pr(E)$
0.002

B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

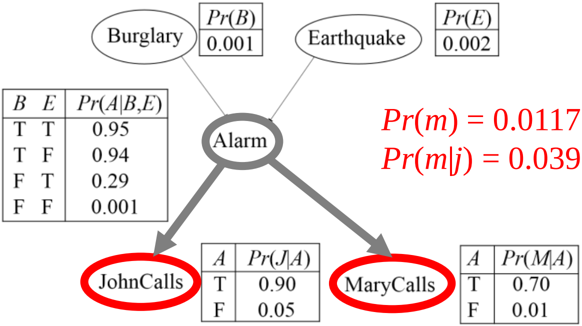
A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

$$Pr(m|a) = Pr(m|a,j) = 0.7$$

贝叶斯网络中的D-分离

- 共同的原因：未给定A的前提下，J与M不独立。

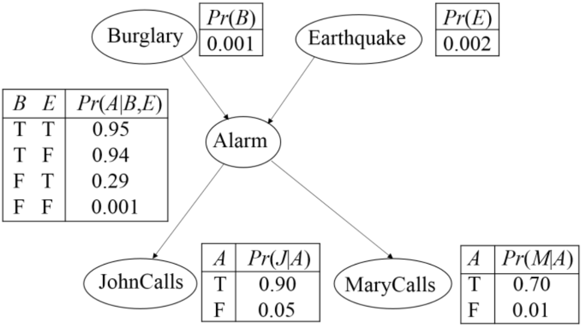


$Pr(m) = 0.011736$

$Pr(m|j) = 0.039972$

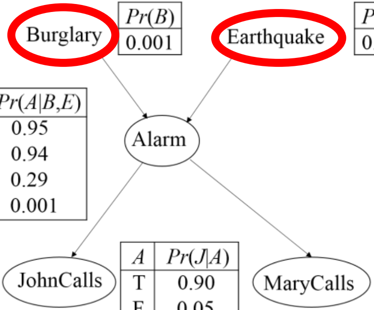
贝叶斯网络中的D-分离

- 共同的作用：



贝叶斯网络中的D-分离

- 共同的作用：



$Pr(B)$
0.001

$Pr(E)$
0.002

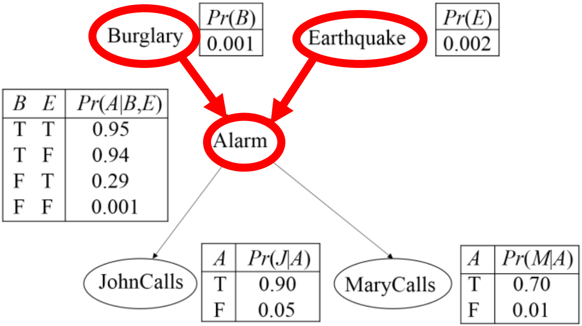
B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

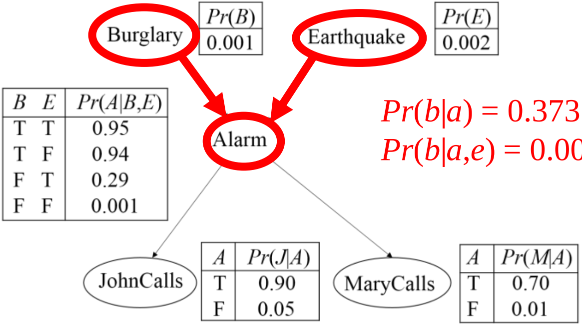
贝叶斯网络中的D-分离

- 共同的作用:



贝叶斯网络中的D-分离

- 共同的作用：给定A的前提下，B与E不是条件独立的。



$Pr(B)$
0.001

$Pr(E)$
0.002

B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

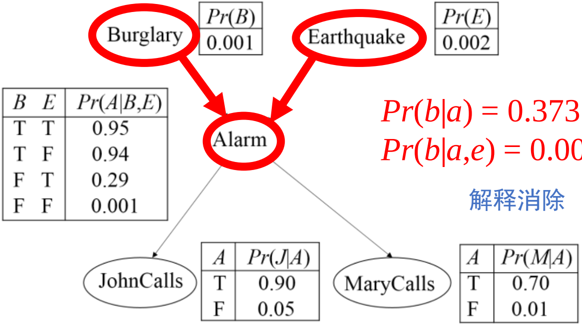
A	$Pr(M\|A)$
T	0.70
F	0.01

$Pr(b|a) = 0.373551$

$Pr(b|a,e) = 0.003268$

贝叶斯网络中的D-分离

- 共同的作用：给定A的前提下，B与E不是条件独立的。



$Pr(B)$
0.001

$Pr(E)$
0.002

B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

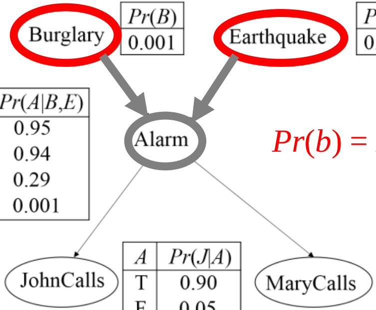
$Pr(b|a) = 0.373551$

$Pr(b|a,e) = 0.003268$

解释消除

贝叶斯网络中的D-分离

- 共同的作用：未给定A的前提下，B与E是独立的。



$Pr(B)$
0.001

$Pr(E)$
0.002

B	E	$Pr(A\|B,E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$Pr(J\|A)$
T	0.90
F	0.05

A	$Pr(M\|A)$
T	0.70
F	0.01

$Pr(b) = Pr(b|e) = 0.001$

贝叶斯网络中的D-分离

- D-分离（Directional separation，D-separation）可用于判断任意两个节点的相关性和独立性。若存在一条路径将这两个节点（直接）连通，则称这两个节点是有向连接（d-connected）的，即这两个节点是相关的；若不存在这样的路径将这两个节点连通，则这两个节点不是有向连接的，则称这两个节点是有向分离的（d-separated），即这两个节点相互独立。

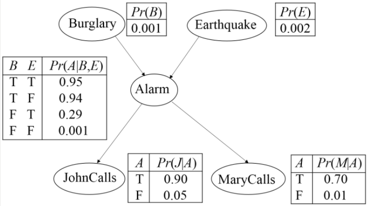
- **定义：**路径p被限定集Z阻塞（block）当且仅当：
  - 路径p含有链结构$A \rightarrow B \rightarrow C$或分连结构$A \leftarrow B \rightarrow C$且中间节点$B$在$Z$中，或者
  - 路径p含有汇连结构$A \rightarrow B \leftarrow C$且汇连节点$B$及其后代都不在$Z$中。

- **定义：**若Z阻塞了节点X和节点Y之间的每一条路径，则称给定Z时，X和Y是D-分离，即给定Z时，X和Y条件独立。

目录

贝叶斯网络的推理

- 给定:
  - 一个贝叶斯网络$Pr(X_1, X_2, \ldots, X_n) = Pr(X_n \mid \mathrm{Par}(X_n)) * Pr(X_{n\text{-}1} \mid \mathrm{Par}(X_{n\text{-}1})) * \ldots * Pr(X_1 \mid \mathrm{Par}(X_1))$
  - 一些证据变量的取值（Evidence）**E** = {a set of values for some of the variables}
- 我们希望推理（计算）如下概率分布：$Pr(X_k \mid \mathbf{E})$，即得到$Pr(X_k = d \mid \mathbf{E})$ for all $d \in \mathrm{Dom}[X_k]$



示例:

给定$Pr(B, E, A, M, J) = Pr(E) * Pr(B) * Pr(A \mid E, B) * Pr(M \mid A) * Pr(J \mid A)$

计算$Pr(B = \mathrm{True} \mid E = \mathrm{False}, M = \mathrm{True}, J = \mathrm{False})$

贝叶斯网络的推理

- 变量消元（Variable Elimination，VE）算法
  - 给定一个贝叶斯网络的一系列条件概率表**F**，查询变量Q，证据变量的取值**E**=e，剩余变量**Z**，需要计算$Pr(Q \mid \mathbf{E})$，则VE算法的流程如下：
  - 对于任意因子$f \in \mathbf{F}$，如果该因子涉及**E**中的一个或多个变量，则将其替换为限制后的因子$f_{\mathbf{E}=e}$；
  - 给定一个消元顺序，对于变量$Z_j \in \mathbf{Z}$：
    - 令$f_1, f_2, \ldots, f_k$为**F**中包含Z_j的所有因子；
    - 将这些因子相乘并在Z_j上求和后得到新的因子：

$$g_j = \sum_{Z_j} f_1 \times f_2 \times \cdots \times f_k\,;$$

    - 从**F**中删掉$f_1, f_2, \ldots, f_k$，并将g_j加入**F**中。
  - 剩下的因子将只涉及查询变量Q，对这些因子相乘并归一化后得到$Pr(Q \mid \mathbf{E})$。

贝叶斯网络的推理

- 因子及相关操作：
  - $f(\mathbf{X}, \mathbf{Y})$指代涉及变量集合$\mathbf{X} \cup \mathbf{Y}$的因子，如条件概率表中的$Pr(C|A)$可以用$f(C, A)$或者$f(A, C)$来表示；
  - 限制：给定$f(X, \mathbf{Y})$且X=a，则$f_{X=a} = h(\mathbf{Y}) = f(a, \mathbf{Y})$;
  - 相乘：给定$f(\mathbf{X}, \mathbf{Y})$和$g(\mathbf{Y}, \mathbf{Z})$，则

$$h(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = f(\mathbf{X}, \mathbf{Y}) \times g(\mathbf{Y}, \mathbf{Z})$$

  - 求和：给定$f(X, \mathbf{Y})$，则该因子在变量X上求和将得到

$$h(\mathbf{Y}) = \sum_{x \in Dom(X)} f(x, \mathbf{Y})$$

f(A,B)		h(B) = $f_{A=a}$	
ab	0.9	b	0.9
a~b	0.1	~b	0.1
~ab	0.4		
~a~b	0.6		

f(A,B)		g(B,C)		h(A,B,C)			
ab	0.9	bc	0.7	abc	0.63	ab~c	0.27
a~b	0.1	b~c	0.3	a~bc	0.08	a~b~c	0.02
~ab	0.4	~bc	0.8	~abc	0.28	~ab~c	0.12
~a~b	0.6	~b~c	0.2	~a~bc	0.48	~a~b~c	0.12

f(A,B)		h(B)	
ab	0.9	b	1.3
a~b	0.1	~b	0.7
~ab	0.4		
~a~b	0.6		

总结

- 其它不确定知识表示和推理方法
  - 确定性理论
    - 该理论由Shortliffe提出，并于1976年首次在血液病诊断专家系统MYCIN中得到了成功应用。
    - 在确定性理论中，不确定性是用可信度来表示的。
  - 证据理论
    - 用于处理不确定性、不精确以及间或不准确的信息。
    - 引入了信任函数来度量不确定性，引用似然函数来处理由于“不知道”引起的不确定性。
  - 模糊逻辑和模糊推理
    - 模糊集合论是1965年由Zadeh提出的，随后，他又将模糊集合论应用于近似或模糊推理，形成了可能性理论。
    - 模糊逻辑可以看作是多值逻辑的扩展。模糊推理是在一组可能不精确的前提下推出一个可能不精确的结论。