机器学习

Machine learning

第九章 概率图模型

Probabilistic Graphical Model

授课人：周晓飞
zhouxiaofei@iie.ac.cn

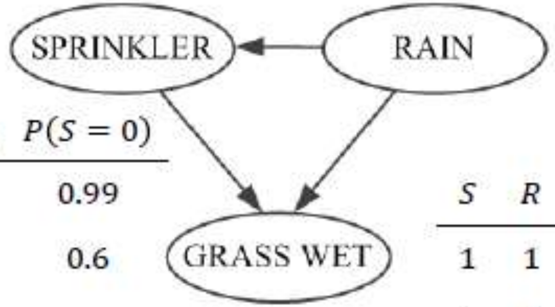
2020-12-10

第九章 概率图模型

学习与推断

基本定义

推断：已知联合概率分布 $P(x_1,x_2,\ldots,x_n)$，估计 $P(\mathbf{x}_Q|\mathbf{x}_E)$，其中 $\mathbf{x}_Q \cap \mathbf{x}_E=\emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E$ 是集合 $\{x_1,x_2,\ldots,x_n\}$的子集。$\mathbf{x}_Q$ 是问题变量，$\mathbf{x}_E$ 是证据变量。



$P(R=1)$	$P(R=0)$
0.2	0.8

R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

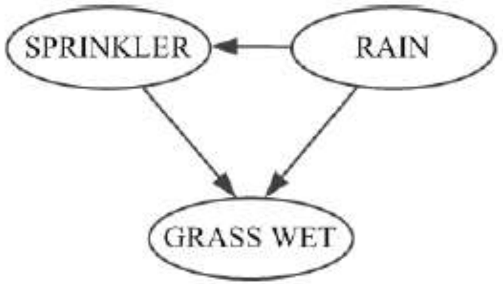
$$P(S=1|G=1)=?$$

$$P(R=1|G=1)=?$$

基本定义

学习：从观测数据 $x^{(1)}, x^{(2)},\ldots,x^{(m)}$中学习联合概率分布 $P(x_1,x_2,\ldots,x_n)$，寻找最符合观测数据的概率图模型。

$$(G,S,R)=(1,1,0)$$
$$(G,S,R)=(1,1,1)$$
$$\vdots$$
$$(G,S,R)=(1,0,1)$$
$$(G,S,R)=(0,0,0)$$



结构学习 (structure learning)

$P(R=1)$	$P(R=0)$
0.2	0.8

R	$P(S=1)$	$P(S=0)$
1	0.01	0.99
0	0.4	0.6

S	R	$P(G=1)$	$P(G=0)$
1	1	0.99	0.01
1	0	0.9	0.1
0	1	0.8	0.2
0	0	0.0	1.0

参数学习 (parameter learning)

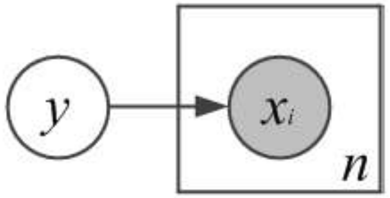
学习与推断

推断

已知联合概率分布 $P(x_1,x_2,\ldots,x_n)$，估计

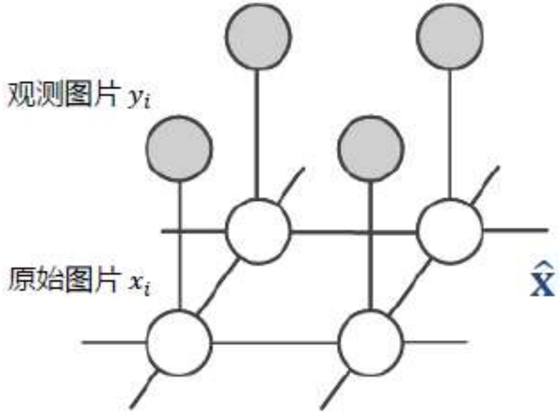
$$P(\mathbf{x}_Q \mid \mathbf{x}_E) = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{P(\mathbf{x}_E)} = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{\sum_{\mathbf{x}_Q} P(\mathbf{x}_Q,\mathbf{x}_E)}$$

其中 $\mathbf{x}_Q \cap \mathbf{x}_E = \emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E = \{ x_1,x_2,\ldots,x_n \}$。



朴素贝叶斯

$\hat{y} = \operatorname{argmax} P(y|\mathbf{x})$



图像去噪

$\hat{\mathbf{x}} = \operatorname{argmax} P(\mathbf{x}|\mathbf{y})$

学习与推断

推断

已知联合概率分布 $P(x_1,x_2,\dots,x_n)$，估计

$$P(\mathbf{x}_Q \mid \mathbf{x}_E) = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{P(\mathbf{x}_E)} = \frac{P(\mathbf{x}_Q,\mathbf{x}_E)}{\boxed{\sum_{\mathbf{x}_Q} P(\mathbf{x}_Q,\mathbf{x}_E)}}$$

核心问题

其中 $\mathbf{x}_Q \cap \mathbf{x}_E = \emptyset$，$\mathbf{x}_Q \cup \mathbf{x}_E = \{x_1,x_2,\dots,x_n\}$。

枚举：假设 $\mathbf{x}_Q$ 有 k 个变量，每个变量的取值个数的期望是 r，则时间复杂度为 r^k。

推断的核心问题：如何高效地计算边际分布 $P(\mathbf{x}_E)=\sum_{x_Q} P(x_Q,x_E)$。

推断方法

精确推断：计算 $P(x_E)$或 $P(x_Q|x_E)$的精确值。

（1）变量消去（variable elimination）。

（2）信念传播（belief propagation）。

近似推断：在较低的时间复杂度下获得原问题的近似解。

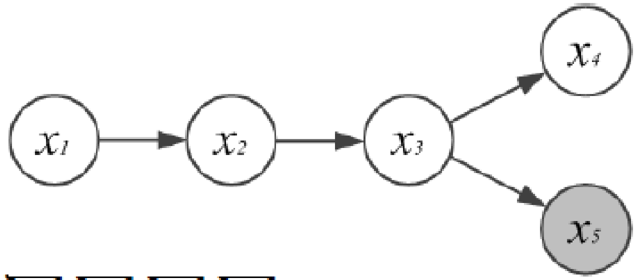
（1）前向采样（forward sampling）。

（2）吉布斯采样（Gibbs sampling）。

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



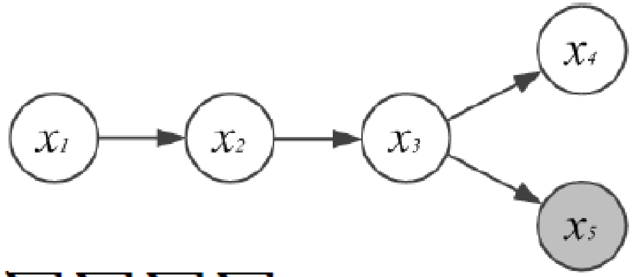
推断 $P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1, x_2, x_3, x_4, x_5)$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



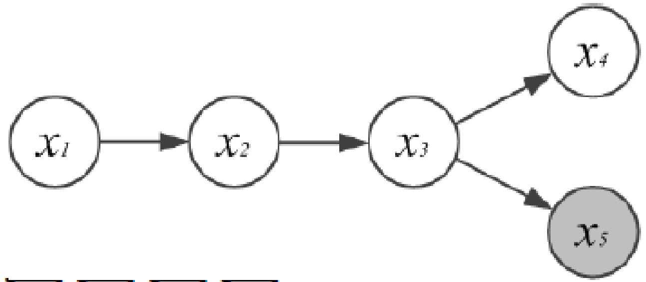
$$P(x_5)=\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1}P(x_1)P(x_2|x_1)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

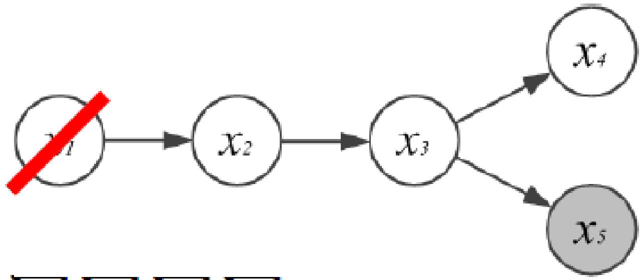
$$= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1)$$

$$= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5)=\sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)$$

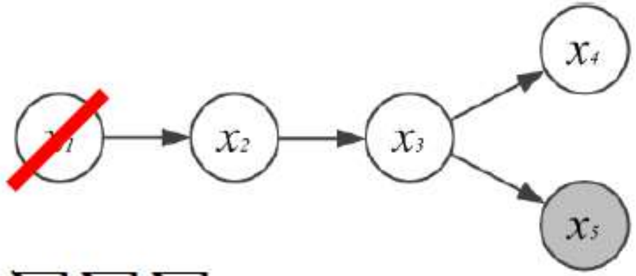
$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1}P(x_1)P(x_2|x_1)$$

$$=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



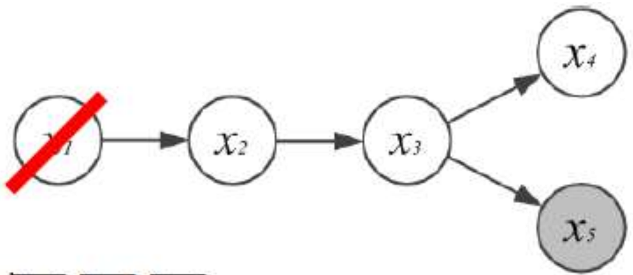
$$P(x_5)=\sum_{x_4}\sum_{x_3}\sum_{x_2}P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2)$$
$$=\sum_{x_4}\sum_{x_3}P(x_4|x_3)P(x_5|x_3)\sum_{x_2}P(x_3|x_2)m_{12}(x_2)$$

学习与推断

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

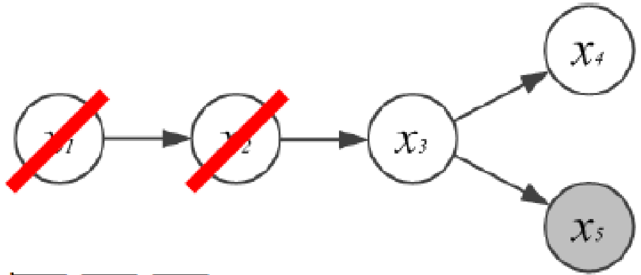


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\
&= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)\sum_{x_2} P(x_3|x_2)m_{12}(x_2) \\
&= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

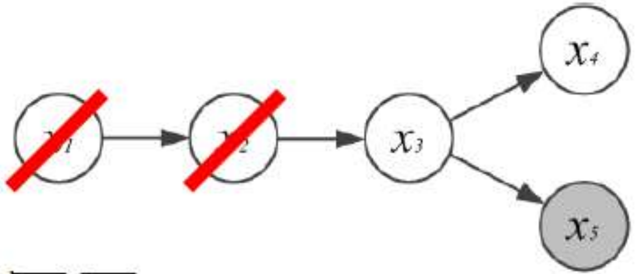


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\
&= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)\sum_{x_2} P(x_3|x_2)m_{12}(x_2) \\
&= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

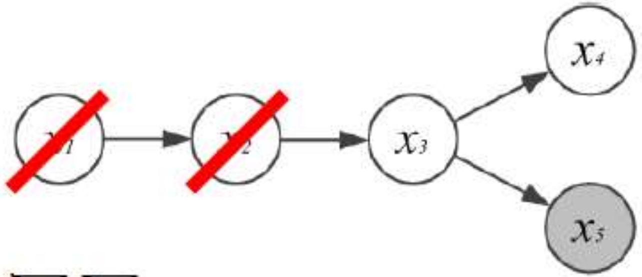


$$P(x_5) = \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$
$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

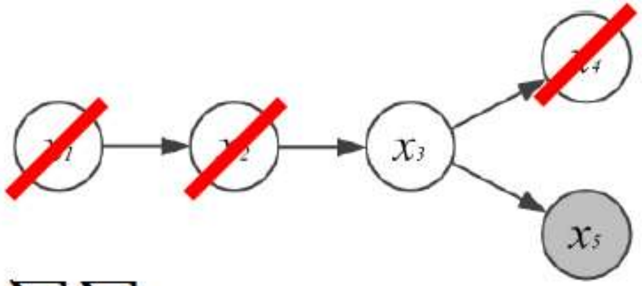


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

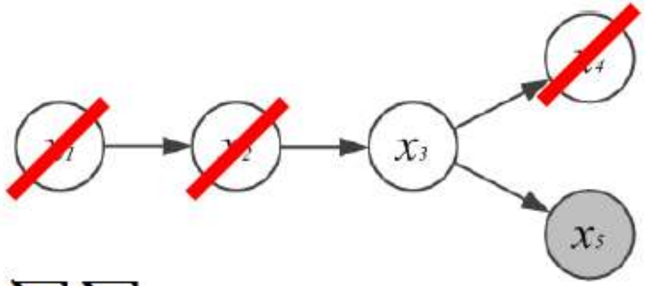


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图

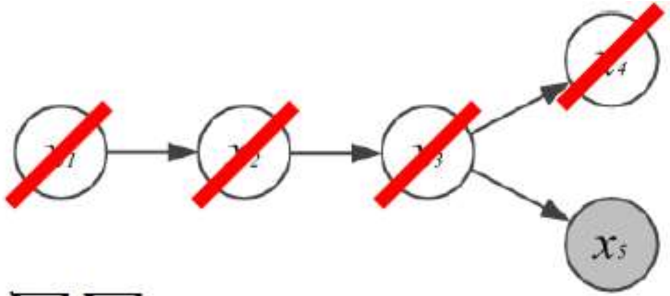


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)\sum_{x_4} P(x_4|x_3) \\
&= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3) = m_{35}(x_5)
\end{aligned}
$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

有向图



$$P(x_5) = \sum_{x_4} \sum_{x_3} P(x_4|x_3)P(x_5|x_3)m_{23}(x_3)$$

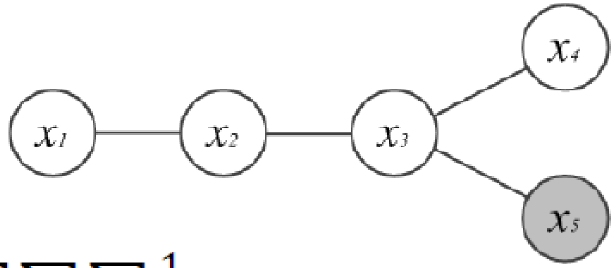
$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3) \sum_{x_4} P(x_4|x_3)$$

$$= \sum_{x_3} P(x_5|x_3)m_{23}(x_3)m_{43}(x_3) = m_{35}(x_5)$$

变量消去

思路：利用图模型的紧凑概率分布形式来削减计算量。

无向图

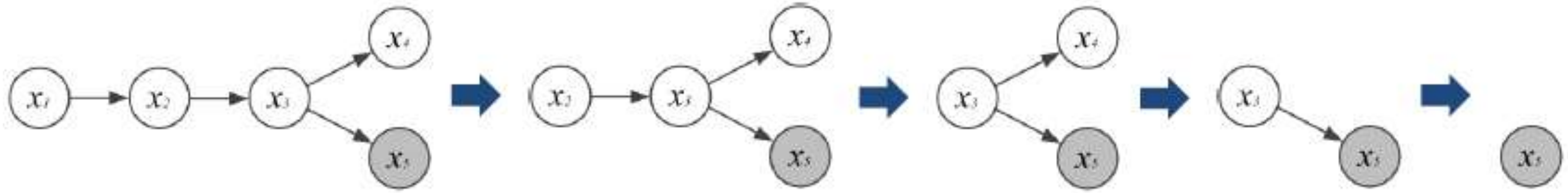


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1}\frac{1}{Z}\psi_{12}(x_1,x_2)\,\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5) \\
&= \sum_{x_4}\sum_{x_3}\sum_{x_2}\frac{1}{Z}\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5)\sum_{x_1}\psi_{12}(x_1,x_2) \\
&= \sum_{x_4}\sum_{x_3}\sum_{x_2}\frac{1}{Z}\psi_{23}(x_2,x_3)\psi_{34}(x_3,x_4)\psi_{35}(x_3,x_5)m_{12}(x_2) = \cdots
\end{aligned}
$$

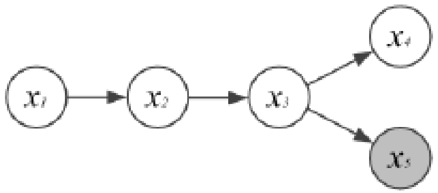
学习与推断

变量消去

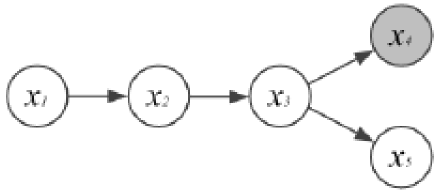
优点：简单直观，代数上的消去对应图中结点的消去。



缺点：针对不同证据变量会造成大量冗余计算。



$$\begin{aligned}P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\\ &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1)\\ &= \cdots\end{aligned}$$

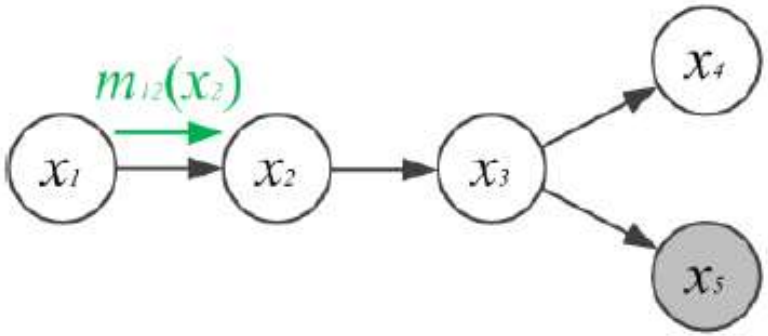


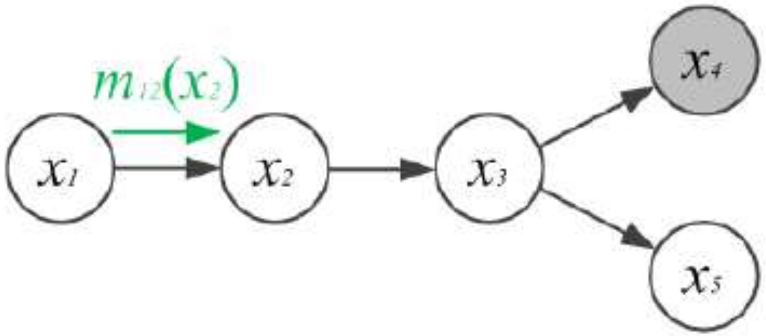
$$\begin{aligned}P(x_4) &= \sum_{x_5}\sum_{x_3}\sum_{x_2}\sum_{x_1} P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\\ &= \sum_{x_5}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1)\\ &= \cdots\end{aligned}$$

学习与推断

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



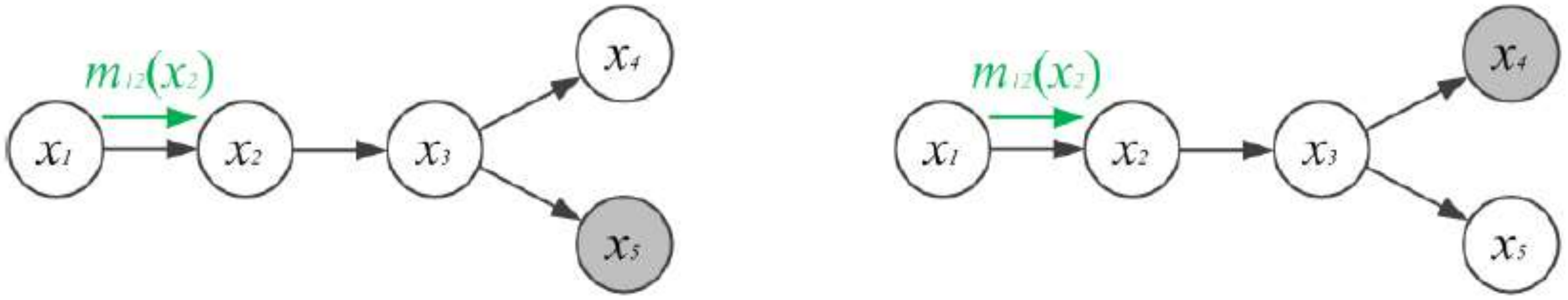


$$
\begin{aligned}
P(x_5) &= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1) \\
&= \sum_{x_4}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\
&= \cdots
\end{aligned}
$$

$$
\begin{aligned}
P(x_4) &= \sum_{x_5}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)\sum_{x_1} P(x_1)P(x_2|x_1) \\
&= \sum_{x_5}\sum_{x_3}\sum_{x_2} P(x_3|x_2)P(x_4|x_3)P(x_5|x_3)m_{12}(x_2) \\
&= \cdots
\end{aligned}
$$

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



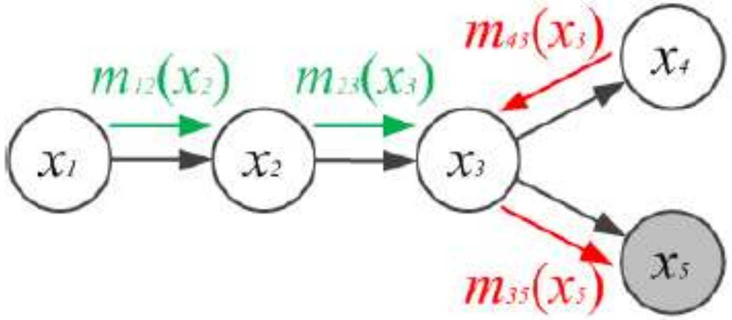
$m_{12}(x_2)$是从 x_1 向 x_2 传递的一条消息。

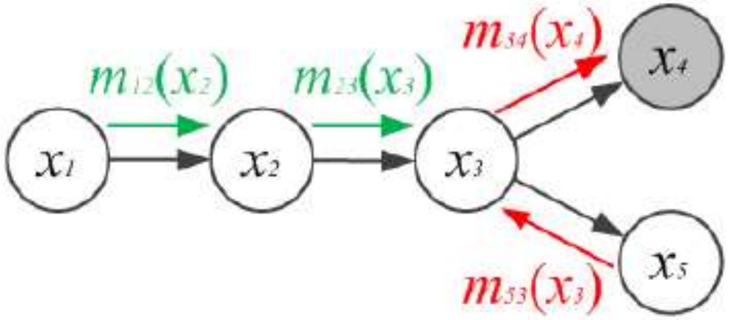
$m_{12}(x_2)$对 x_1 进行了求和，是关于 x_2 的函数。

$m_{12}(x_2)$仅与图的拓扑结构有关，与证据变量的选取无关（可复用）。

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



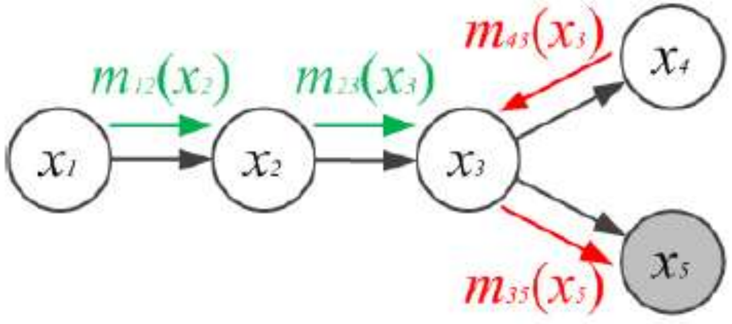


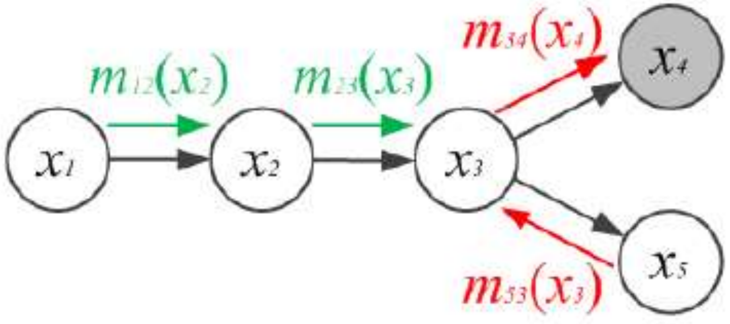
$m_{12}(x_2)$，$m_{23}(x_3)$可重复使用。

$m_{43}(x_3)$，$m_{35}(x_5)$，$m_{34}(x_4)$，$m_{53}(x_3)$需重新计算。

信念传播

思路：将变量消去过程中产生的中间结果视为可复用的消息，避免重复计算。



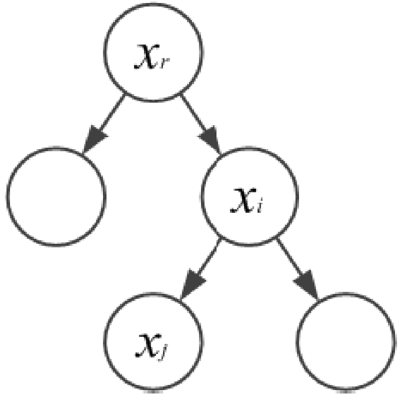


消息传递仅在邻接变量之间发生。

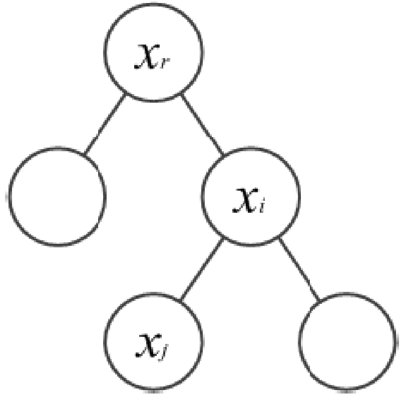
消息传递与边的方向性无关。

信念传播

树结构：有向树=无向树



$$P(\mathbf{x}) = P(x_r) \prod_{(i,j)\in E} P(x_j | x_i)$$

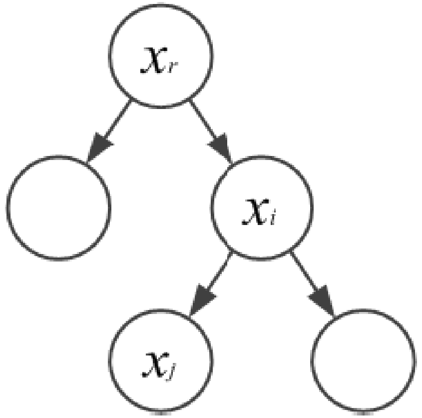


$$P(\mathbf{x}) = \frac{1}{Z} \prod_{i\in V} \psi(x_i) \prod_{(i,j)\in E} \psi(x_i, x_j)$$

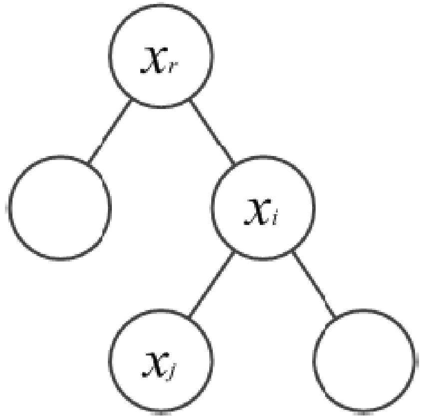
学习与推断

信念传播

树结构：有向树=无向树



=



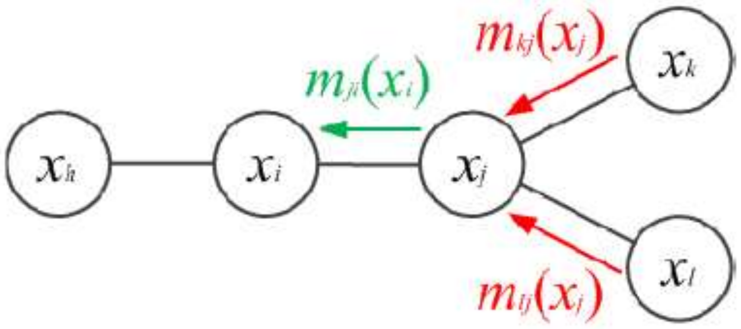
$$P(\mathbf{x}) = P(x_r)\prod_{(i,j)\in E} P(x_j|x_i)$$

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i\in V}\psi(x_i)\prod_{(i,j)\in E}\psi(x_i,x_j)$$

$$\boxed{Z = 1;\ \psi(x_r) = P(x_r);\ \psi(x_{\neg r}) = 1;\ \psi(x_i,x_j) = P(x_j|x_i)}$$

信念传播

树结构上的消息传递：



消息计算公式： （N(j)表示 j 的邻域节点）

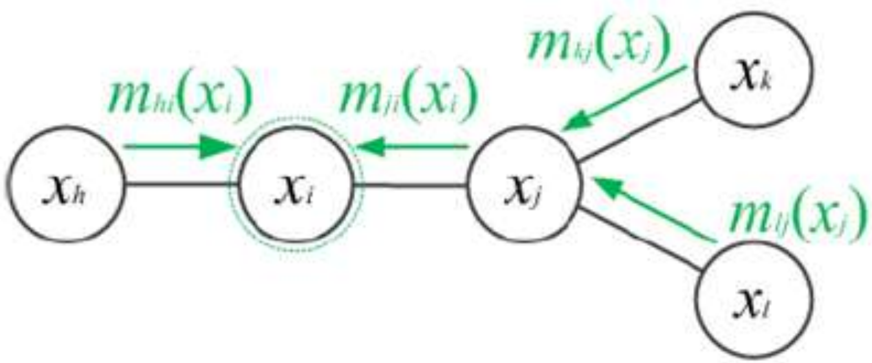
$$m_{ji}(x_i) = \sum_{x_j} \psi(x_i, x_j) \prod_{k \in N(j)\backslash i} m_{kj}(x_j)$$

信念传播

边际分布：

$$P(x_i) \propto \prod_{k \in N(i)} m_{ki}(x_i)$$

结点 k 对结点 i 的信念

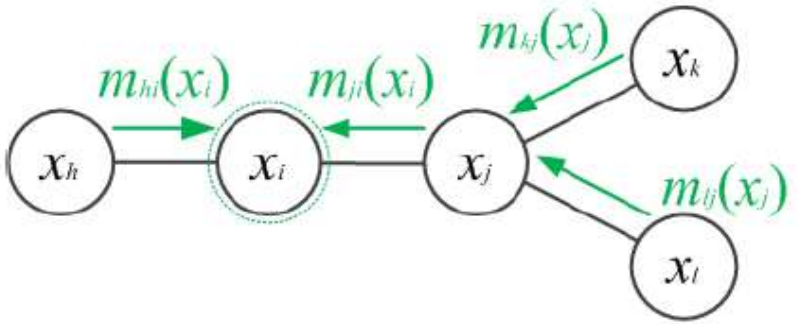


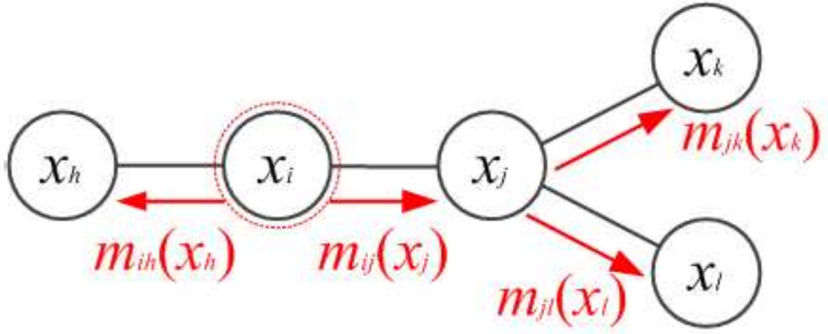
信念传播

二次扫描算法：

（1） 指定一个根结点，从所有叶结点开始向根节点传递消息，直到根结点收到所有邻接结点的消息。

（2） 从根结点开始向叶结点传递消息，直到所有叶结点均收到消息。





第九章 概率图模型

近似推断

推断方法

精确推断：

变量消去、信念传播。

计算复杂度随着极大团规模的增长呈指数级增长，适用范围有限。

近似推断：

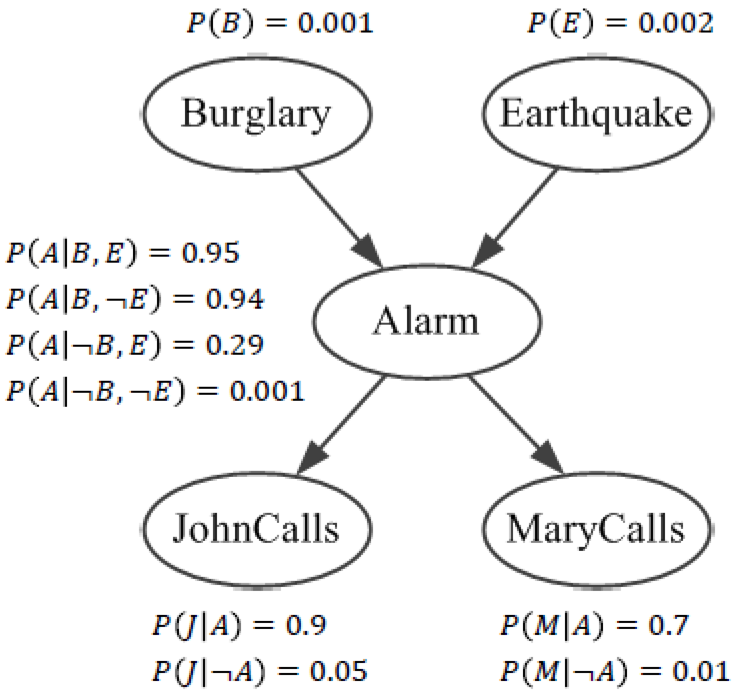
前向采样、吉布斯采样。

通过采样一组服从特定分布的样本，来近似原始分布，适用范围更广，操作性更强。

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。

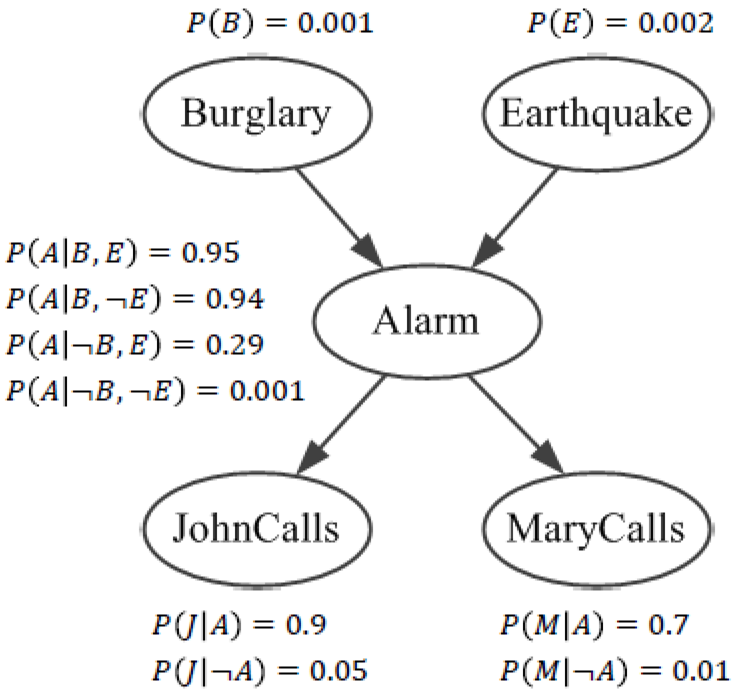


$$P(B=1 \mid E=0, J=1) = ?$$

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。



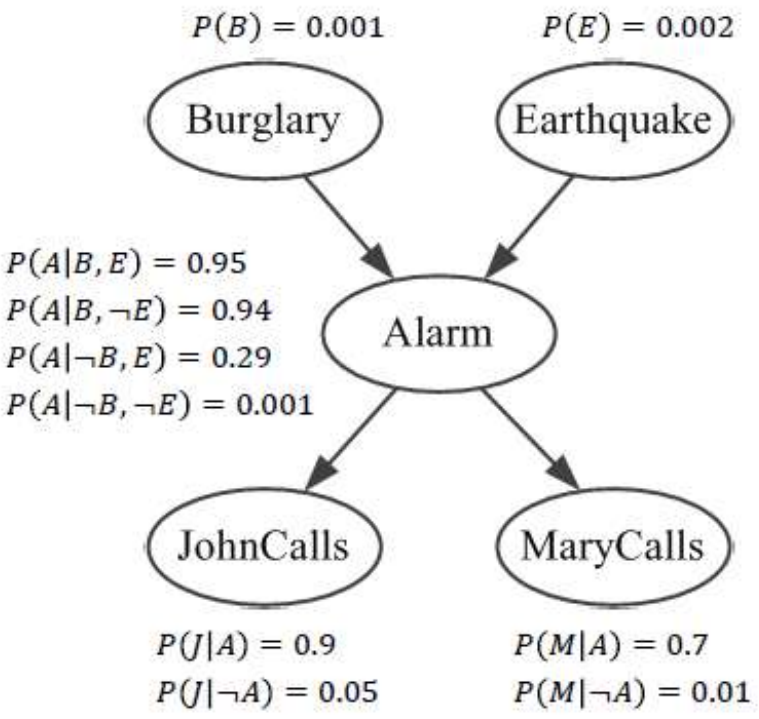
$$P(B=1|E=0,J=1)=?$$

$$\begin{aligned}
&P(B=1|E=0,J=1)\\
&=\frac{P(B=1,E=0,J=1)}{P(E=0,J=1)}\\
&=\frac{\sum_{A,M}P(B=1,E=0,A,J=1,M)}{\sum_{B,A,M}P(B,E=0,A,J=1,M)}\\
&=\cdots
\end{aligned}$$

采样后，进行需要的概率统计！

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。

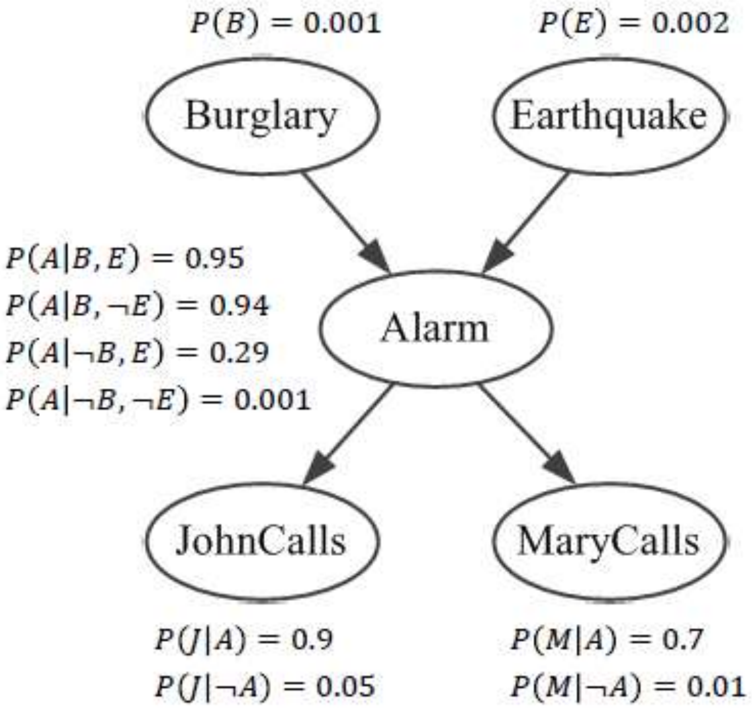


- 采样 $B \sim \mathrm{Ber}(0.001)$
  - 采样 $r \sim \mathrm{U}(0,1)$：若$r < 0.001$，$B = 1$；否则 $B = 0$

近似推断

前向采样

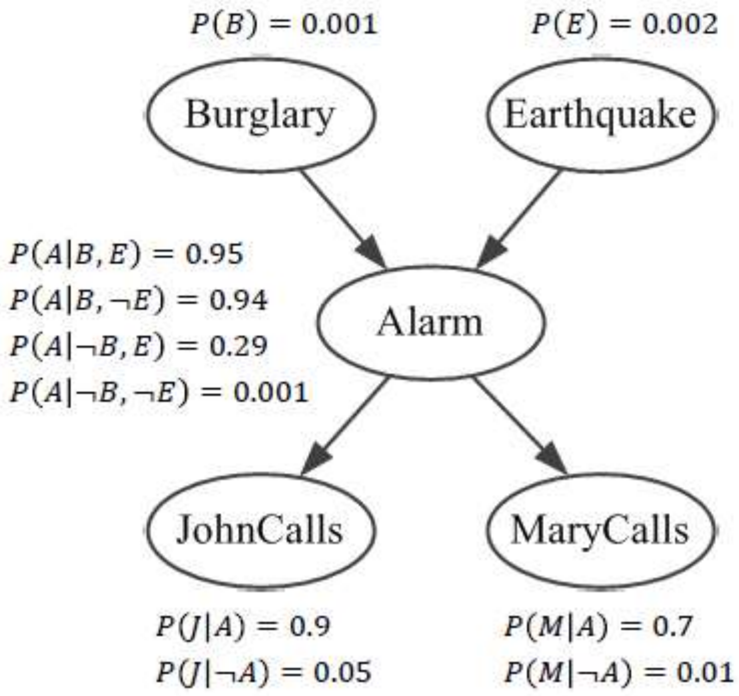
思路：依据贝叶斯网络的（条件）概率直接采样。



- 采样 $B \sim \mathrm{Ber}(0.001)$
- 采样 $E \sim \mathrm{Ber}(0.002)$

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。



- 采样 $B \sim \mathrm{Ber}(0.001)$
- 采样 $E \sim \mathrm{Ber}(0.002)$
- 若 $B = 1, E = 1$，采样 $A \sim \mathrm{Ber}(0.95)$

  若 $B = 1, E = 0$，采样 $A \sim \mathrm{Ber}(0.94)$

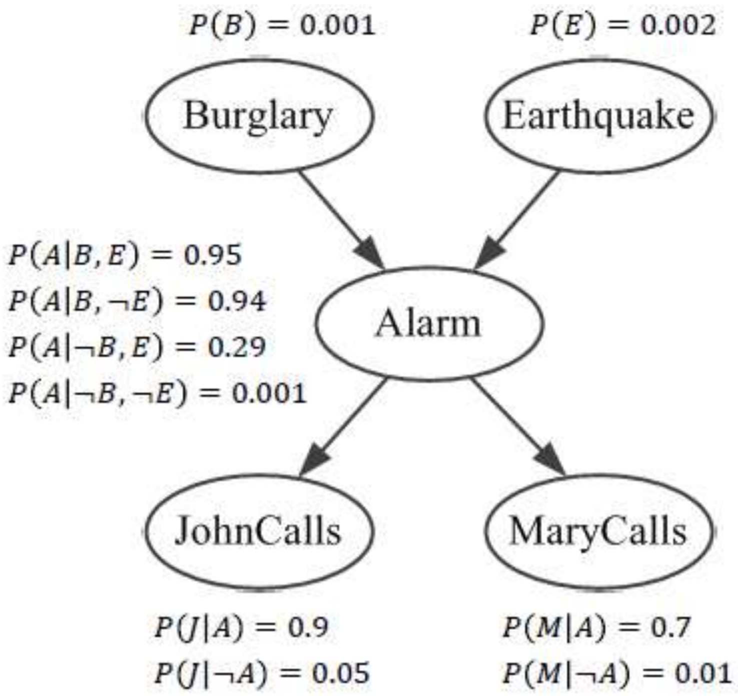
  若 $B = 0, E = 1$，采样 $A \sim \mathrm{Ber}(0.29)$

  若 $B = 0, E = 0$，采样 $A \sim \mathrm{Ber}(0.001)$

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。



- 采样 $B \sim \mathrm{Ber}(0.001)$
- 采样 $E \sim \mathrm{Ber}(0.002)$
- 若 $B = 1, E = 1$，采样 $A \sim \mathrm{Ber}(0.95)$

  若 $B = 1, E = 0$，采样 $A \sim \mathrm{Ber}(0.94)$

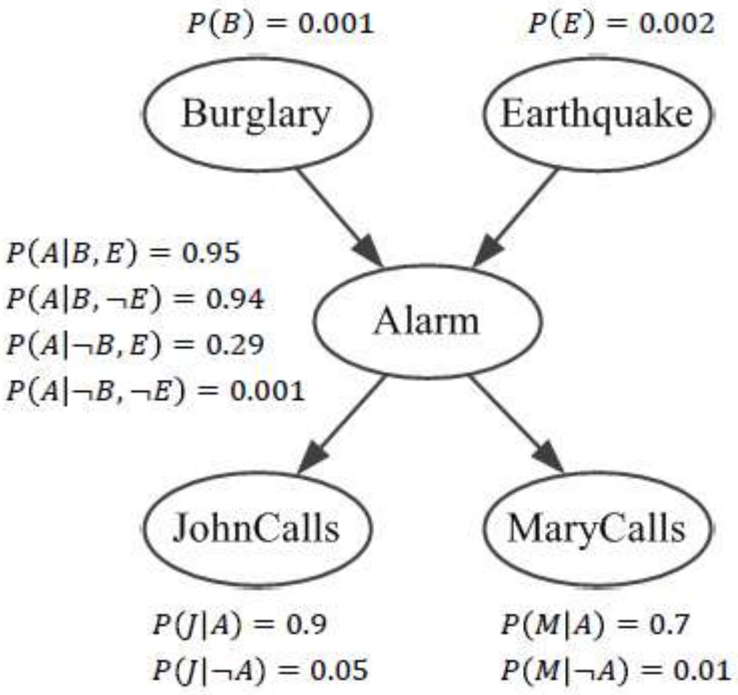
  若 $B = 0, E = 1$，采样 $A \sim \mathrm{Ber}(0.29)$

  若 $B = 0, E = 0$，采样 $A \sim \mathrm{Ber}(0.001)$
- 若 $A = 1$，采样 $J \sim \mathrm{Ber}(0.9)$；否则采样 $J \sim \mathrm{Ber}(0.05)$

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。



- 采样 $B \sim \mathrm{Ber}(0.001)$
- 采样 $E \sim \mathrm{Ber}(0.002)$
- 若 $B = 1, E = 1$，采样 $A \sim \mathrm{Ber}(0.95)$

  若 $B = 1, E = 0$，采样 $A \sim \mathrm{Ber}(0.94)$

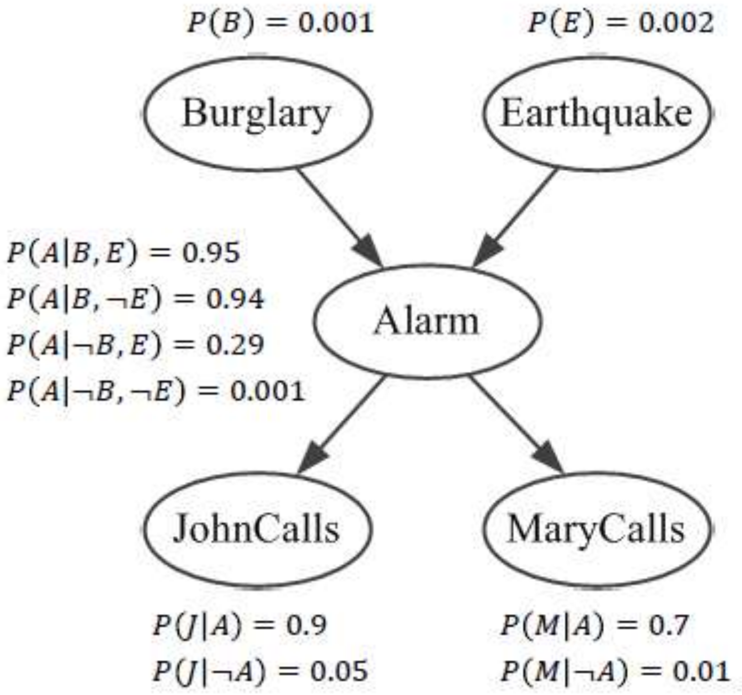
  若 $B = 0, E = 1$，采样 $A \sim \mathrm{Ber}(0.29)$

  若 $B = 0, E = 0$，采样 $A \sim \mathrm{Ber}(0.001)$
- 若 $A = 1$，采样 $J \sim \mathrm{Ber}(0.9)$；否则采样 $J \sim \mathrm{Ber}(0.05)$
- 若 $A = 1$，采样 $M \sim \mathrm{Ber}(0.7)$；否则采样 $M \sim \mathrm{Ber}(0.01)$

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。

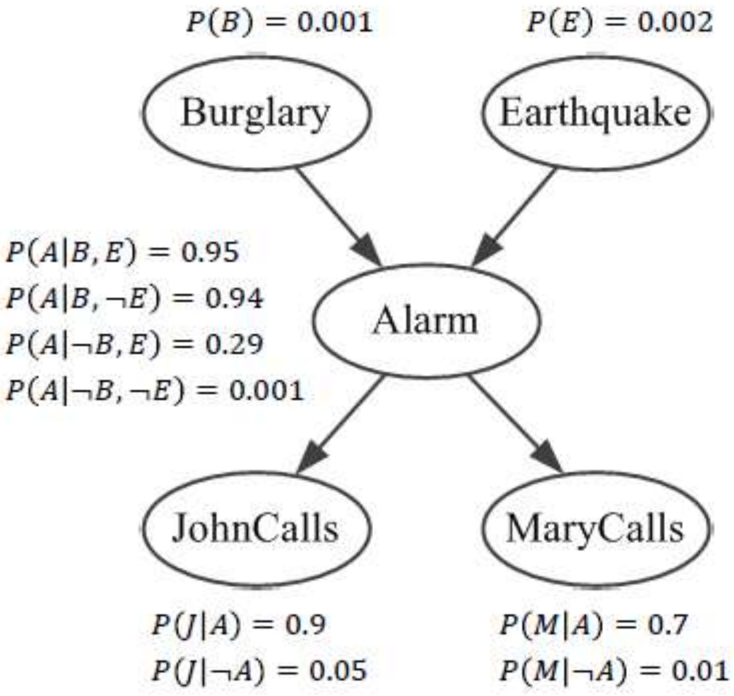


B	E	A	J	M
0	0	0	0	0
1	0	1	1	0
0	0	0	0	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
1	0	1	1	1
0	0	0	0	0
0	0	1	1	0
0	0	0	0	0

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。



B	E	A	J	M
0	0	0	0	0
1	0	1	1	0
0	0	0	0	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
1	0	1	1	1
0	0	0	0	0
0	0	1	1	0
0	0	0	0	0

$$P(B=1|E=0,J=1) = {}^{2}/_{3}$$

近似推断

前向采样

思路：依据贝叶斯网络的（条件）概率直接采样。

缺点：I) 对于小概率事件采样困难，可能经过很多次采样也无法获得足够多的样本；

II) 仅适用于贝叶斯网络，不适用于马尔可夫随机场。

近似推断

吉布斯采样

思路：直接依照条件概率 $P(\mathbf{x}_Q|\mathbf{x}_E)$采样。

Procedure Gibbs Sampling

Fix $\mathbf{x}_E$ and randomly initialize other variables $\mathbf{x}_O := \{x_1, \cdots, x_n\} \backslash \mathbf{x}_E$

repeat

    foreach $x_i \in \mathbf{x}_O$ **do**

        Sample $x \sim P(x_i | x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ /* x_i 在其他所有变量当前取值下的条件概率 */

        $x_i \leftarrow x$

until Convergence

return The later samples

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。

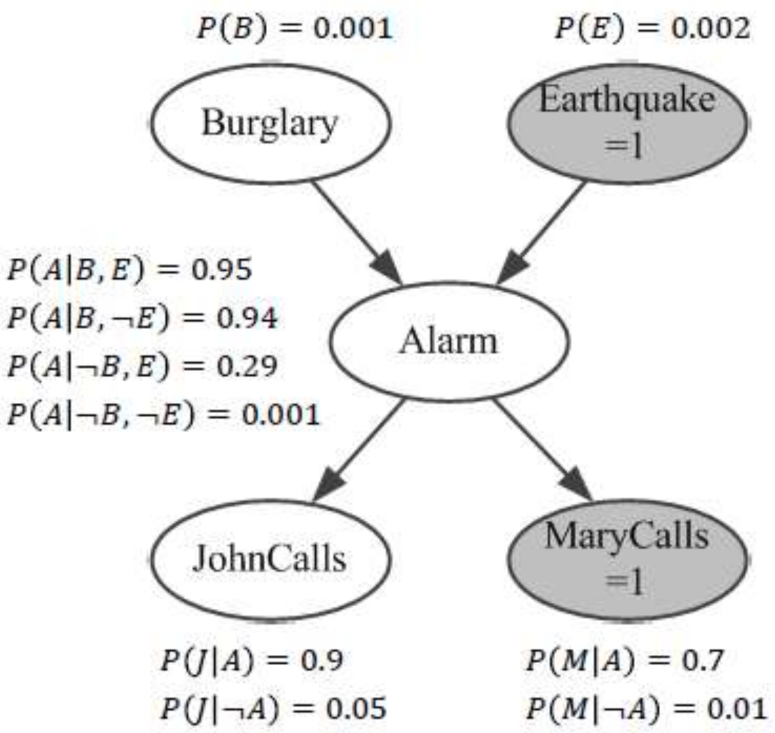
马尔可夫毯的性质：

$$P(x_i \mid x_1, \ldots x_{i-1}, x_{i-1}, \ldots, x_n) = P(x_i \mid MB(x_i))$$

近似推断

吉布斯采样

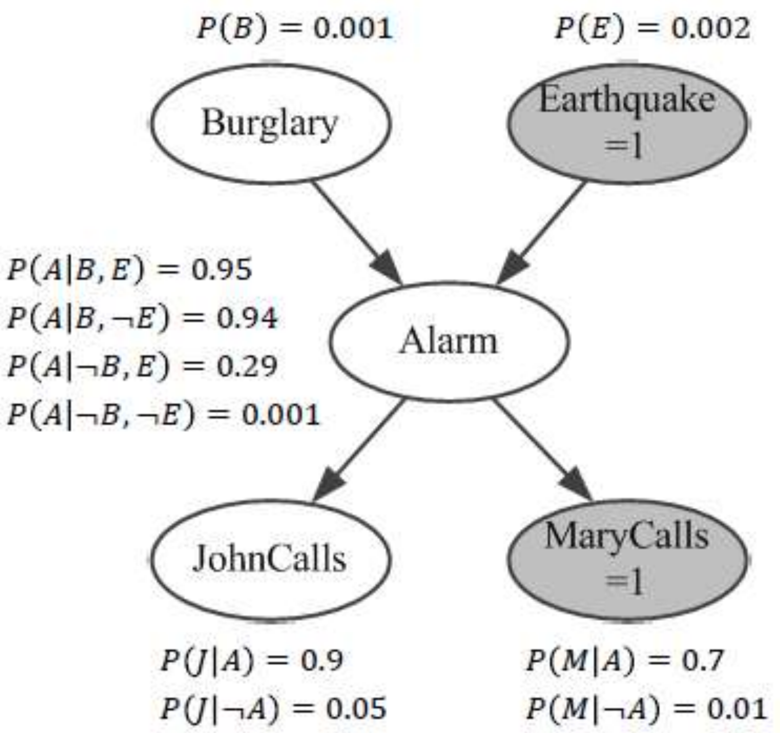
思路：直接依照条件概率 $P(x_Q|x_E)$采样。



$$P(B=1|E=1,M=1)=?$$

吉布斯采样

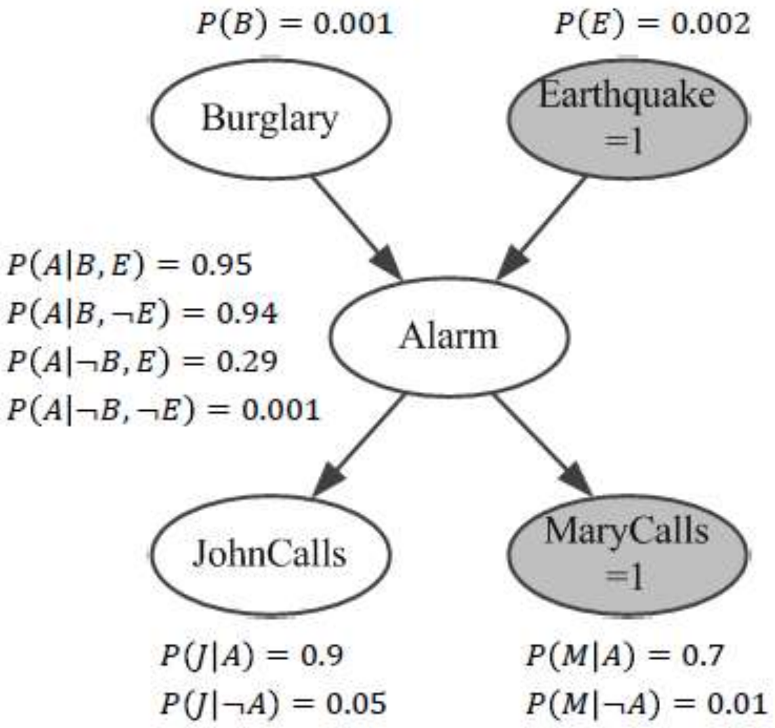
思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$					

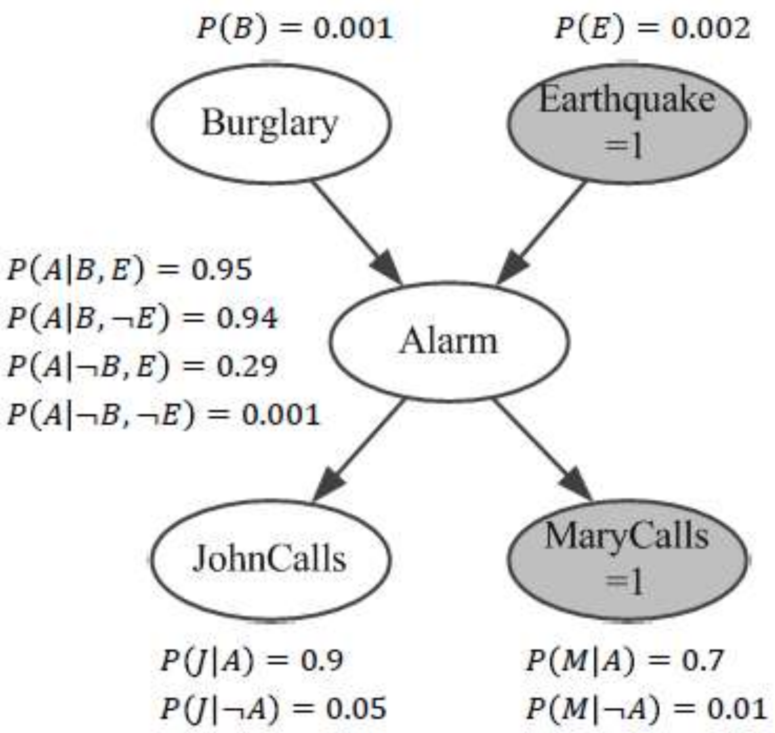
- 采样 B，马尔可夫毯 $\{A, E\}$

$$P(B|\neg A, E) \propto P(B)P(\neg A|B, E) = 0.001 \times 0.05$$

$$P(\neg B|\neg A, E) \propto P(\neg B)P(\neg A|\neg B, E) = 0.999 \times 0.71$$

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0				

- 采样 B，马尔可夫毯 $\{A, E\}$

  $P(B|\neg A, E) \propto P(B)P(\neg A|B,E) = 0.001 \times 0.05$

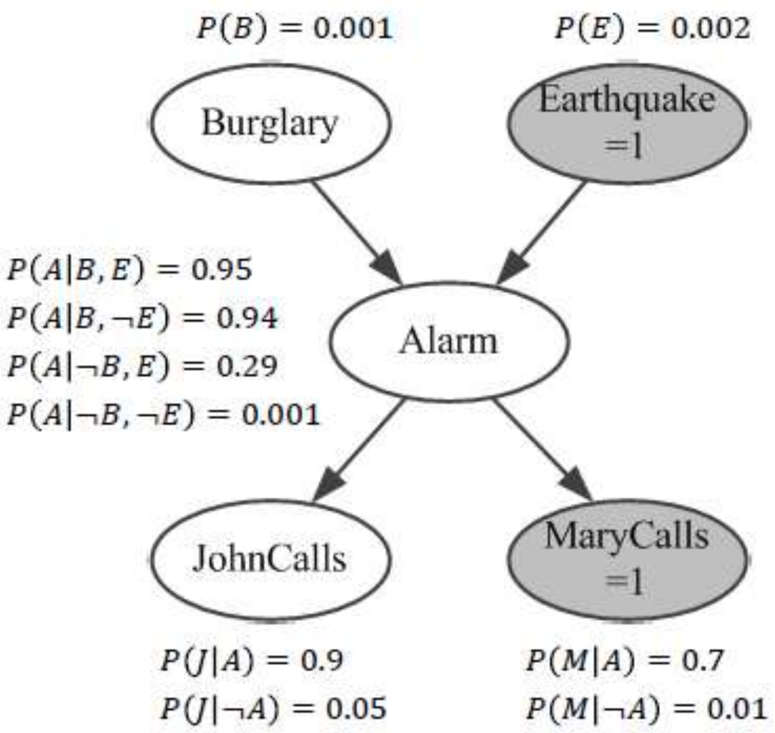
  $P(\neg B|\neg A, E) \propto P(\neg B)P(\neg A|\neg B,E) = 0.999 \times 0.71$

- 采得 $B = 0$

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。

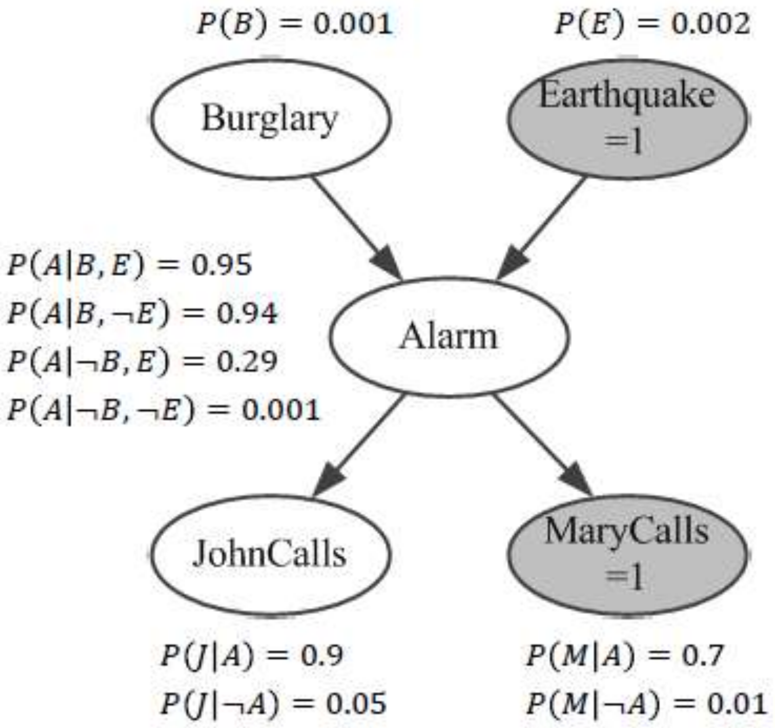


Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0	1			

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t = 0$	0	1	0	0	1
$t = 1$	0	1			

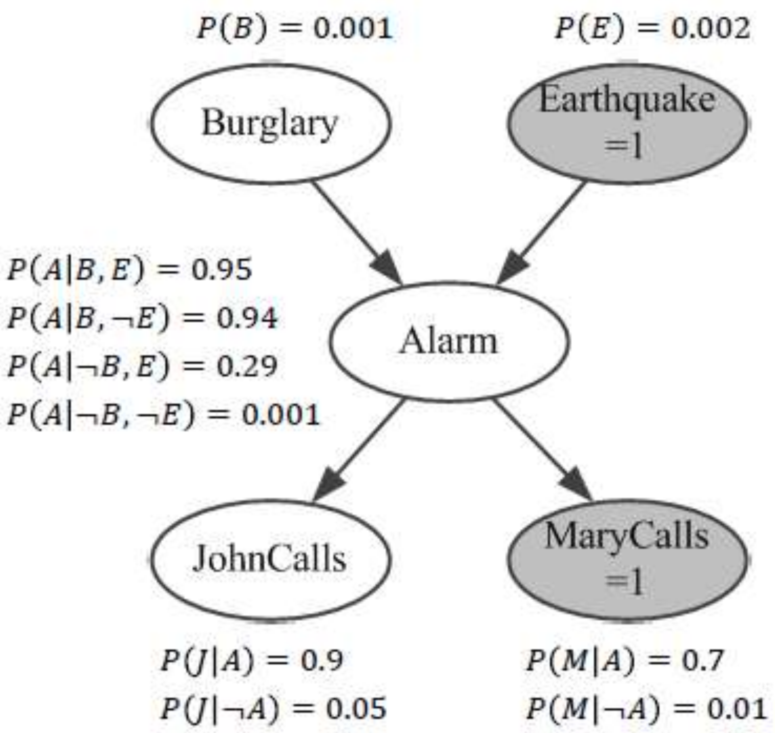
- 采样 A，马尔可夫毯 $\{B, E, J, M\}$

$$P(A|\neg B, E, \neg J, M) \propto P(A|\neg B, E)P(\neg J|A)P(M|A) = 0.29 \times 0.1 \times 0.7$$

$$P(\neg A|\neg B, E, \neg J, M) \propto P(\neg A|\neg B, E)P(\neg J|\neg A)P(M|\neg A) = 0.71 \times 0.95 \times 0.01$$

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0	1	1		

- 采样 A，马尔可夫毯 $\{B, E, J, M\}$

$$P(A|\neg B, E, \neg J, M) \propto P(A|\neg B, E)P(\neg J|A)P(M|A) = 0.29 \times 0.1 \times 0.7$$

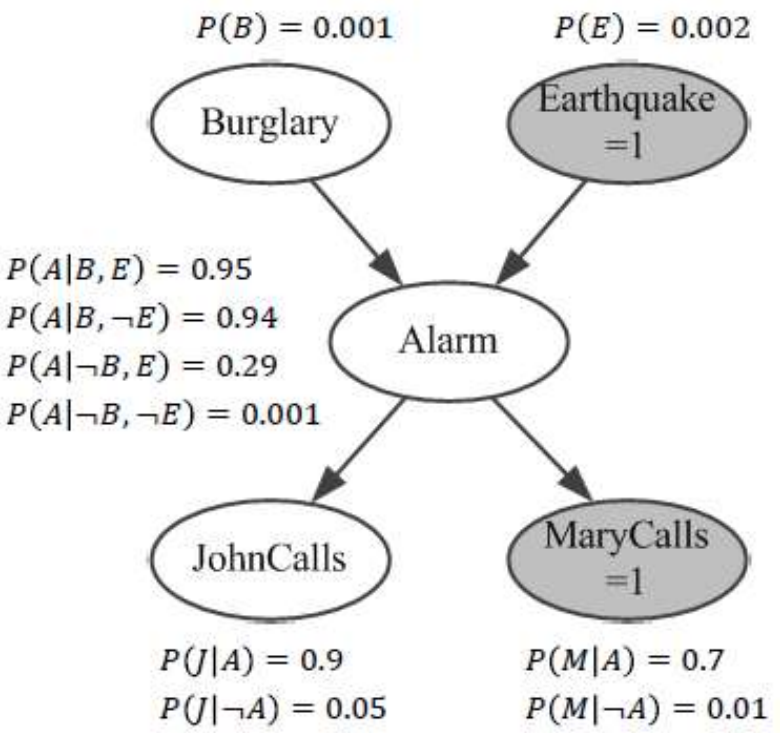
$$P(\neg A|\neg B, E, \neg J, M) \propto P(\neg A|\neg B, E)P(\neg J|\neg A)P(M|\neg A) = 0.71 \times 0.95 \times 0.01$$

- 采得 $A = 1$

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0	1	1		

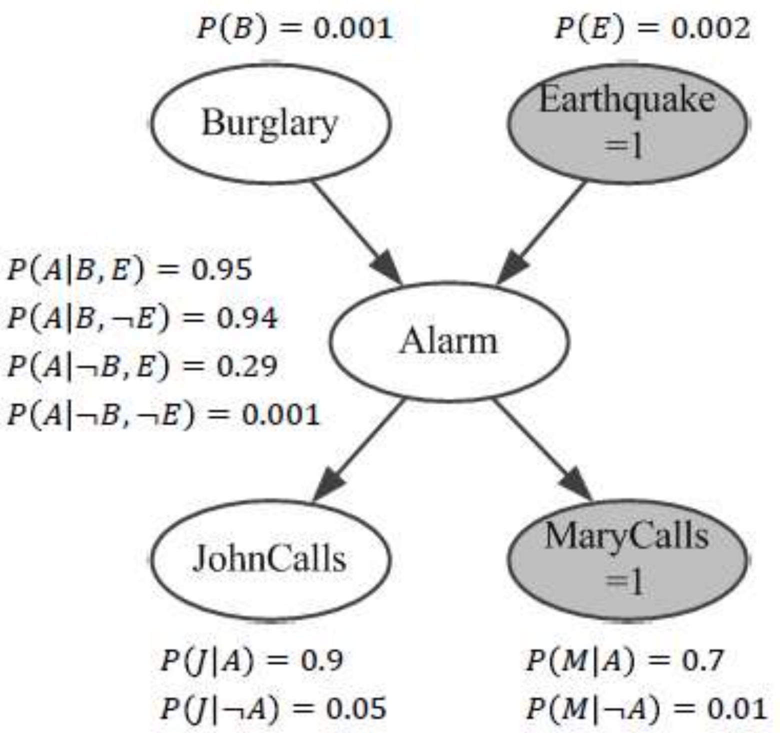
- 采样 J，马尔可夫毯 $\{A\}$

$P(J|A) = 0.9$

$P(\neg J|A) = 0.1$

吉布斯采样

思路：直接依照条件概率 P(x_Q|x_E)采样。



Iteration	B	E	A	J	M
$t = 0$	0	1	0	0	1
$t = 1$	0	1	1	1	

- 采样 J，马尔可夫毯 $\{A\}$

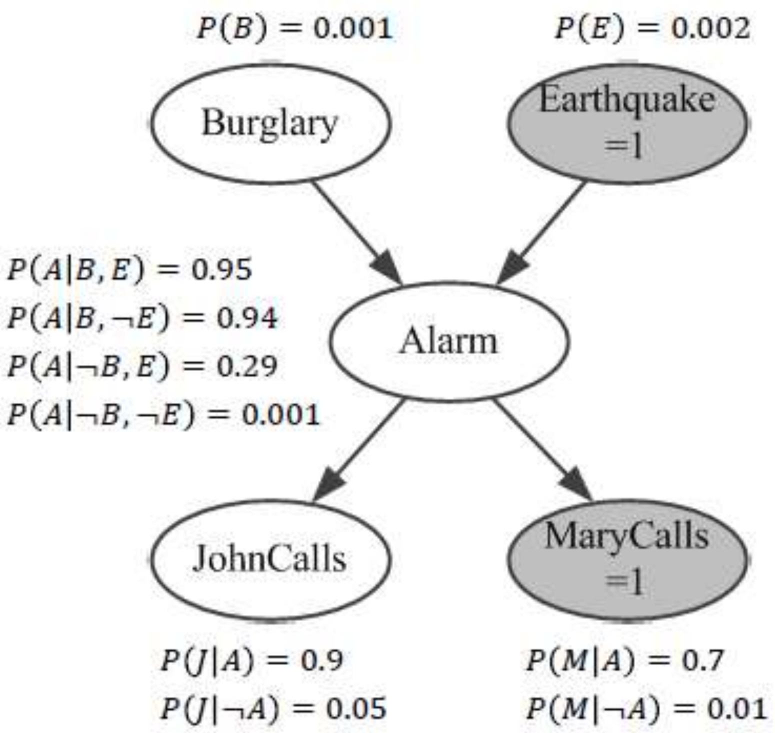
  $P(J|A) = 0.9$

  $P(\neg J|A) = 0.1$

- 采得 $J = 1$

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。

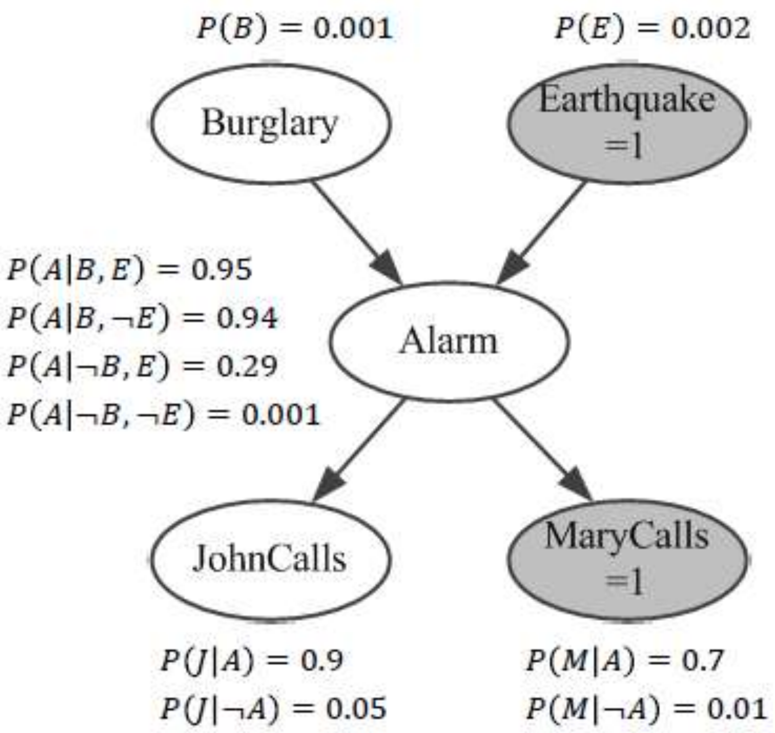


Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0	1	1	1	1

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。



Iteration	B	E	A	J	M
$t=0$	0	1	0	0	1
$t=1$	0	1	1	1	1
$t=2$	0	1	0	1	1
$t=3$	1	1	1	1	1

近似推断

吉布斯采样

思路：直接依照条件概率 $P(x_Q|x_E)$采样。

优点：I) 直接从 $P(x_Q|x_E)$采样，解决小概率事件采样难的问题；（ Fixed X_E ）

II) 同时适用于贝叶斯网络和马尔可夫随机场；

III) 简单易推导，时间复杂度低。

本讲参考文献

1. 《统计机器学习--第九章：概率图模型》课件，王泉，国科大网络安全学院，2017。

致谢王泉副研究员！感谢王泉提供了《概率图模型》课件供本章教学参考！