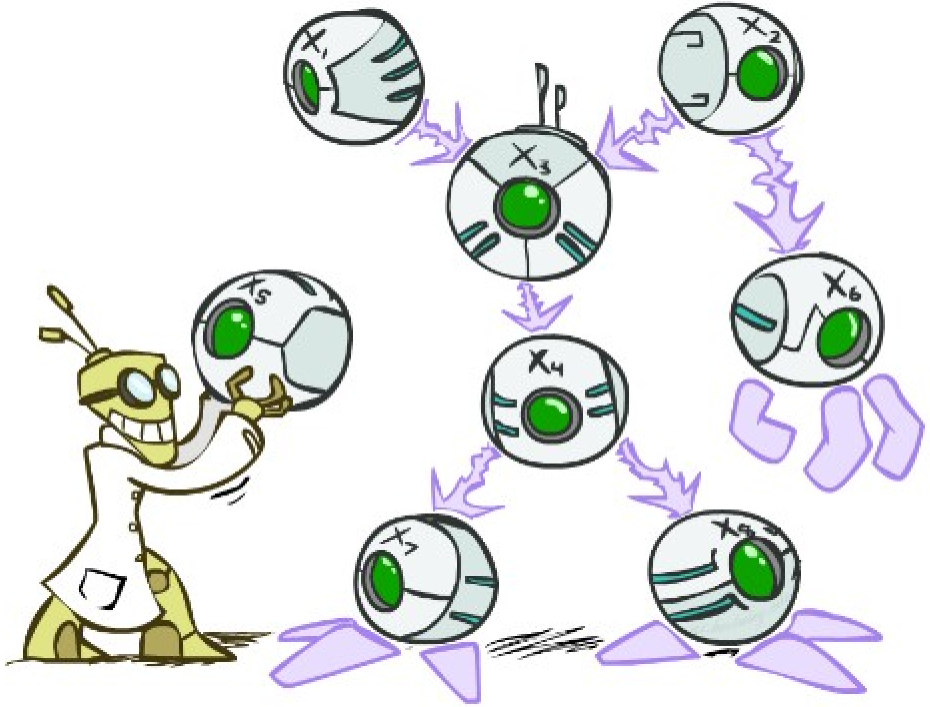
不确定性知识的表示与推理

第十四章 概率推理



materials are available at http://ai.berkeley.edu

提纲

- 第十四章 概率推理
  - 14.1 独立性与条件独立性
  - 14.2 不确定问题的知识表示 - 贝叶斯网络
  - 14.3 贝叶斯网络的语义
  - 14.4 精确推理：枚举推理、 变量消元

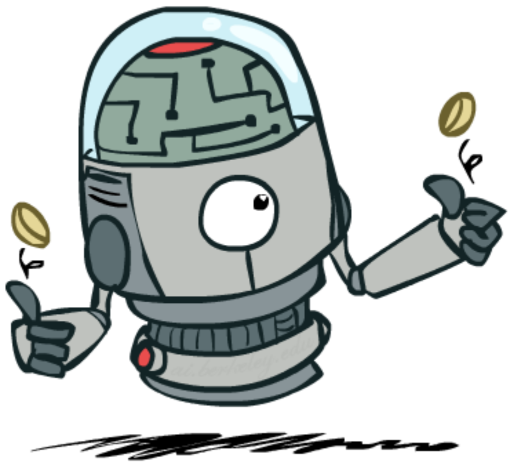
独立性

- 引入第四个变量 *Weather* ，有四个取值 <*sunny, rainy, cloudy, snow*>

- 完全联合分布：**P**(*Toothache, Catch, Cavity, Weather*) ，有 2 × 2 × 2 × 4 = 32 个条目．

$$P(toothache, catch, cavity, cloudy)$$
$$= P(cloudy \mid toothache, catch, cavity) P(toothache, catch, cavity)$$

$$P(cloudy \mid toothache, catch, cavity) = P(cloudy)$$



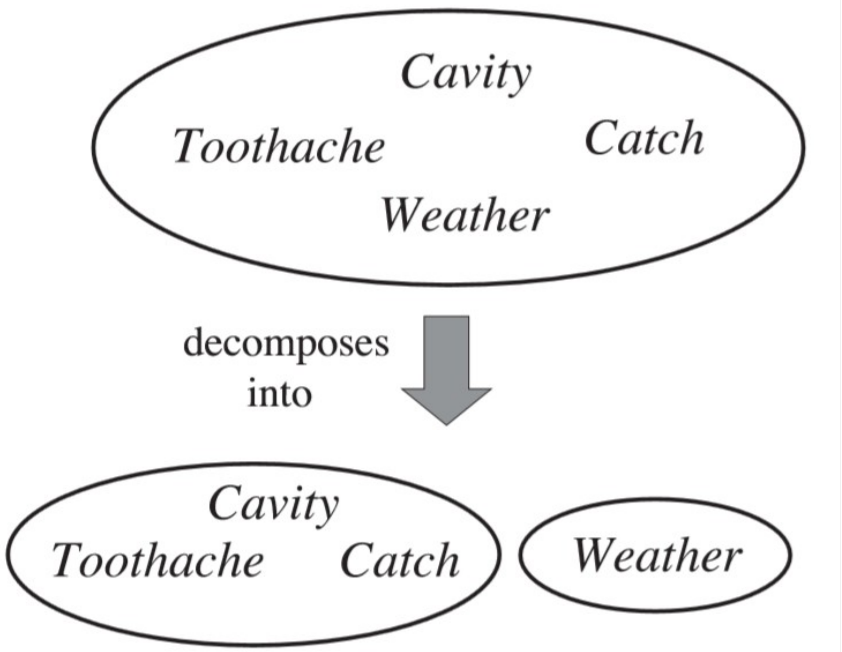
- $P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)$

独立性

- 写出通用的联合分布：

P(*Toothache*, *Catch*, *Cavity*, *Weather*)

= **P**(*Toothache*, *Catch*, *Cavity*) **P**(*Weather*)



- *Weather* 与其它三个变量之间相互独立

- 完全 联合分布表中的 32 (8x4) 个条目可以降低为 12(8+4) 个

独立性

- 两个随机变量 *A* 和 *B* 之间**独立**，当且仅当：

$$\mathbf{P}(A, B) = \mathrm{P}(A)\,\mathrm{P}(B) \quad \text{或} \quad \mathbf{P}(A \mid B) = \mathbf{P}(A) \quad \text{或} \quad \mathbf{P}(B \mid A) = \mathbf{P}(B)$$

  - 独立性断言有助于减小问题域表示，并降低推理复杂度
  - e.g., 可以假定 *Toothache* 和 *Weather* 之间相互独立

条件独立性

绝对独立性是强大的，但现实应用中很少。领域知识通常具有成百个变量，它们之间并不完全独立

- 条件独立性：
  - 给定随机变量 C，两个随机变量 A 和 B 是条件独立的，当且仅当：

$$\mathbf{P}(A, B \mid C) = \mathbf{P}(A \mid C)\, P(B \mid C)$$

或 $$\mathbf{P}(A \mid B, C) = \mathbf{P}(A \mid C)$$

或 $$\mathbf{P}(B \mid A, C) = \mathbf{P}(B \mid C)$$

条件独立性

- 问题域：
  - Traffic
  - Umbrella
  - Raining

条件独立性

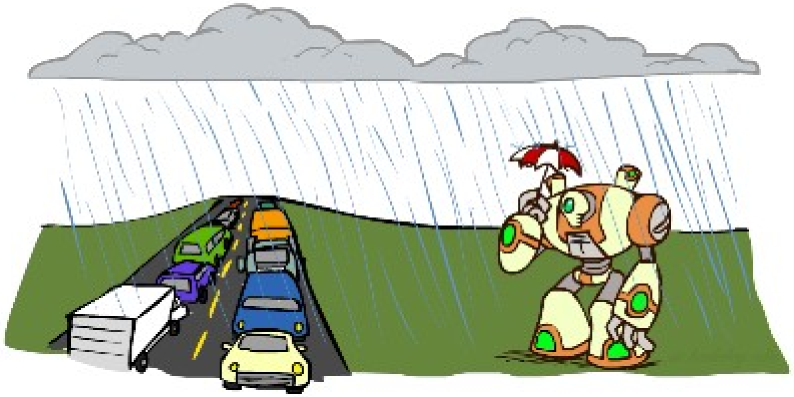
- 链式法则： $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- 计算完全联合概率分布

P(*Rain, Traffic, Umbrella*)

= **P**(*Rain*) **P**(*Traffic* | *Rain*) <u>**P**(*Umbrella*| *Rain, Traffic*)</u>　（链式法则）

= **P**(*Rain*) **P**(*Traffic* | *Rain*) **P**(*Umbrella*| *Rain*)　（条件独立的假设）



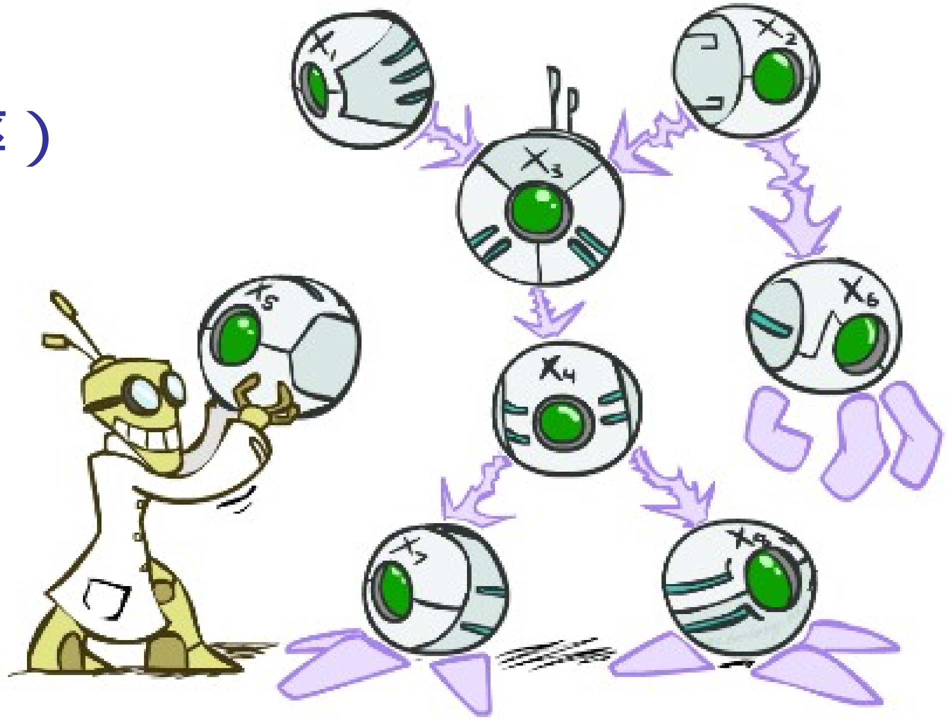
- 条件独立假设表示的方法：贝叶斯网络

提纲

- 第十四章 概率推理
  - 14.1 独立性与条件独立性
  - 14.2 不确定问题的知识表示 - 贝叶斯网络
  - 14.3 贝叶斯网络的语义
  - 14.4 精确推理：枚举推理、 变量消元

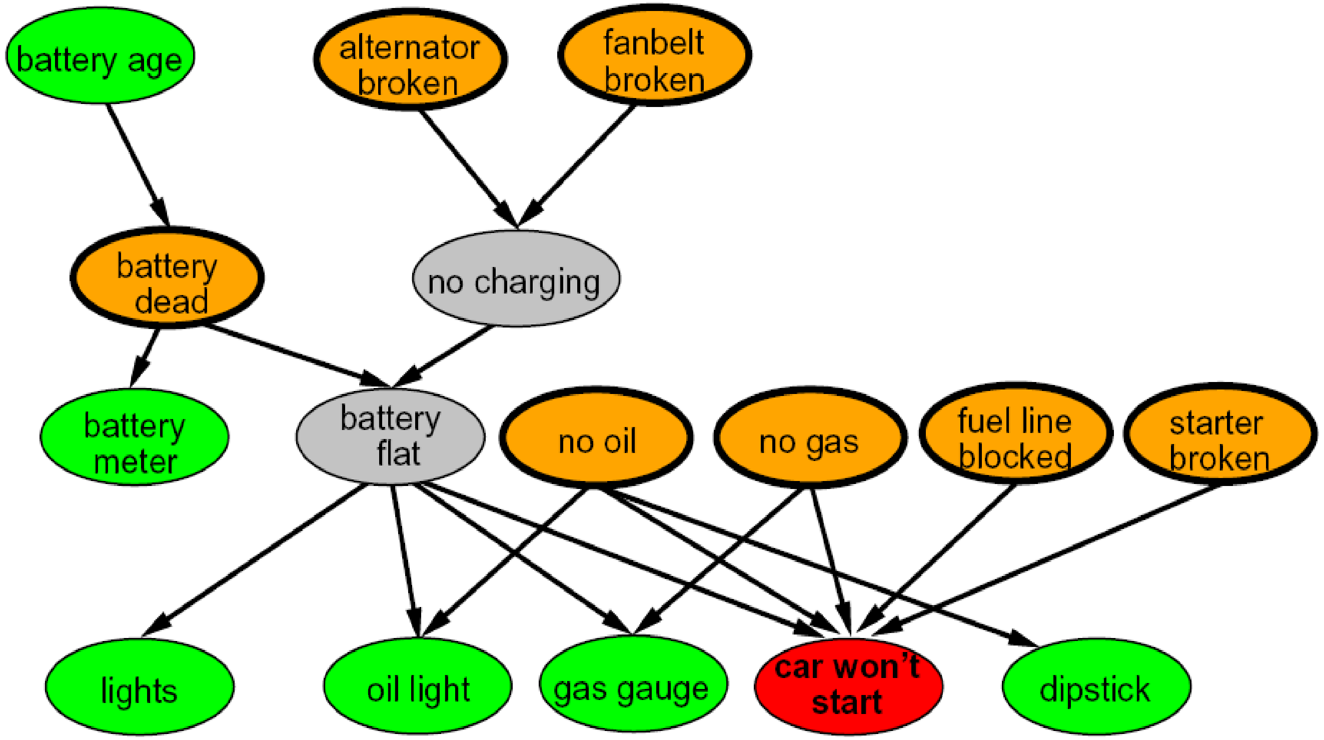
贝叶斯网络

- 完全概率分布能够回答关于问题域的任何问题，但是：
  - 随着变量数目增多会增大到不可操作的程度
  - 为每个可能世界逐个指定概率是不符合实际的
- 贝叶斯网络：一种使用简单局部分布（条件概率）描述复杂联合分布的技术
  - 是一个有向无环图，也称之为图模型
  - 用于表示变量之间的依赖关系
  - 本质上，表示任何完全联合概率分布



贝叶斯网络示例：Car Diagnosis

- 贝叶斯网络：
  - 某个领域的概率模型

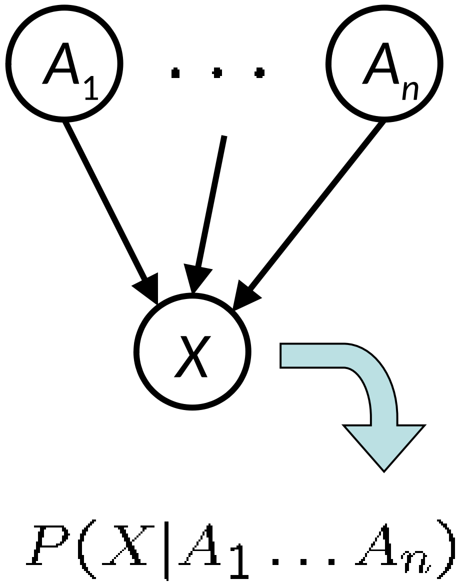


- Questions：
  - 推理：如何计算 P（ X | e ）？
  - 表示法：可以编码什么样的分布？
  - 建模：什么 BN 最适合给定的领域？

贝叶斯网络模型

- 语法：
  - 结点对应随机变量 $X_i, A_1,\ldots, A_n$ 可以是已知的或者是待观测的
  - 有向边表示连接的变量间的"直接影响"
  - 条件概率分布，量化父结点对该结点 X_i 的影响

$$\mathbf{P}(X_i \mid Parents(X_i))$$

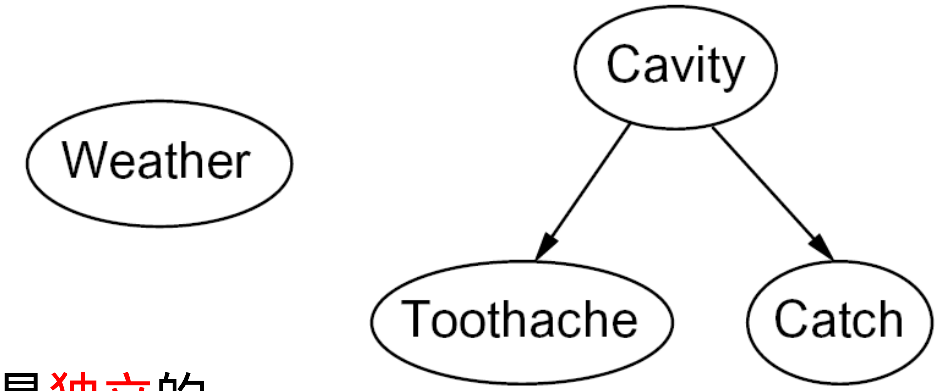


- 条件概率表 (conditional probability table, CPT)，为每个变量指定其相对于父结点的条件概率

贝叶斯网络 = 拓扑结构（图）+ 条件概率表

贝叶斯网络模型

- 贝叶斯网络的拓扑结构
  - 用一种精确简洁的方式描述了在问题域中成立的独立和条件独立关系

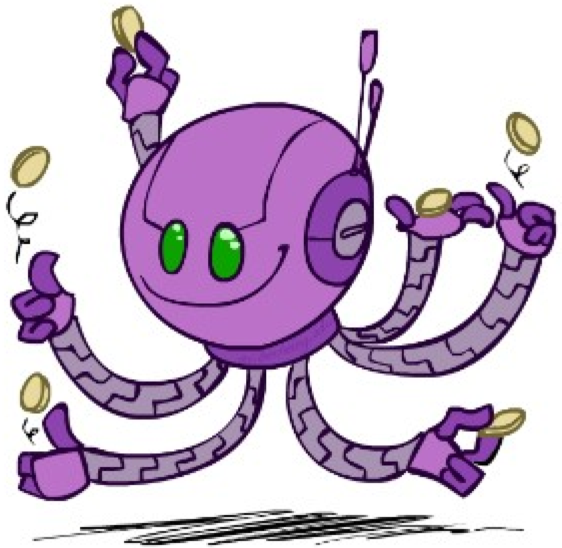


  - *Weather* 与其他变量是独立的
  - 给定 *Cavity* 条件下，*Toothache* 和 *Catch* 是条件独立的

示例 1：抛硬币

- N 次独立抛硬币

X_1 X_2 $\cdots$ X_n



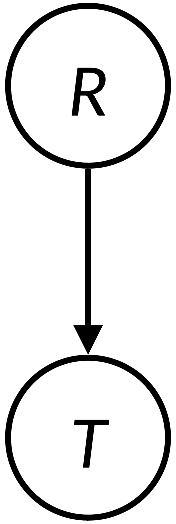
- 变量之间没有关系：绝对独立性

示例 2: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

- Model 2: rain causes traffic （因果方向）

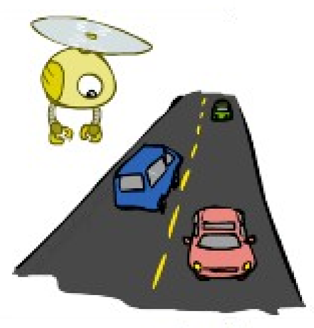


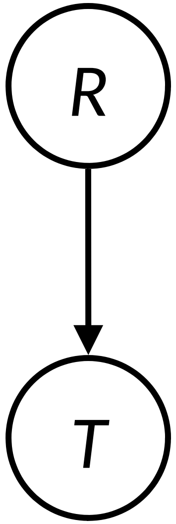
- Agent using model 2 better

示例 2: Traffic

- 贝叶斯网络的拓扑结构 + CPT

Model 2: rain causes traffic（因果方向）





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

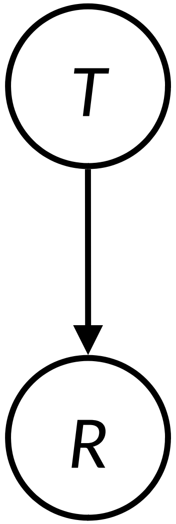
$P(T,R)$

+r	+t	
+r	-t	
-r	+t	
-r	-t	

示例 2: Reverse Traffic

- 逆因果方向？

Model 3:



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

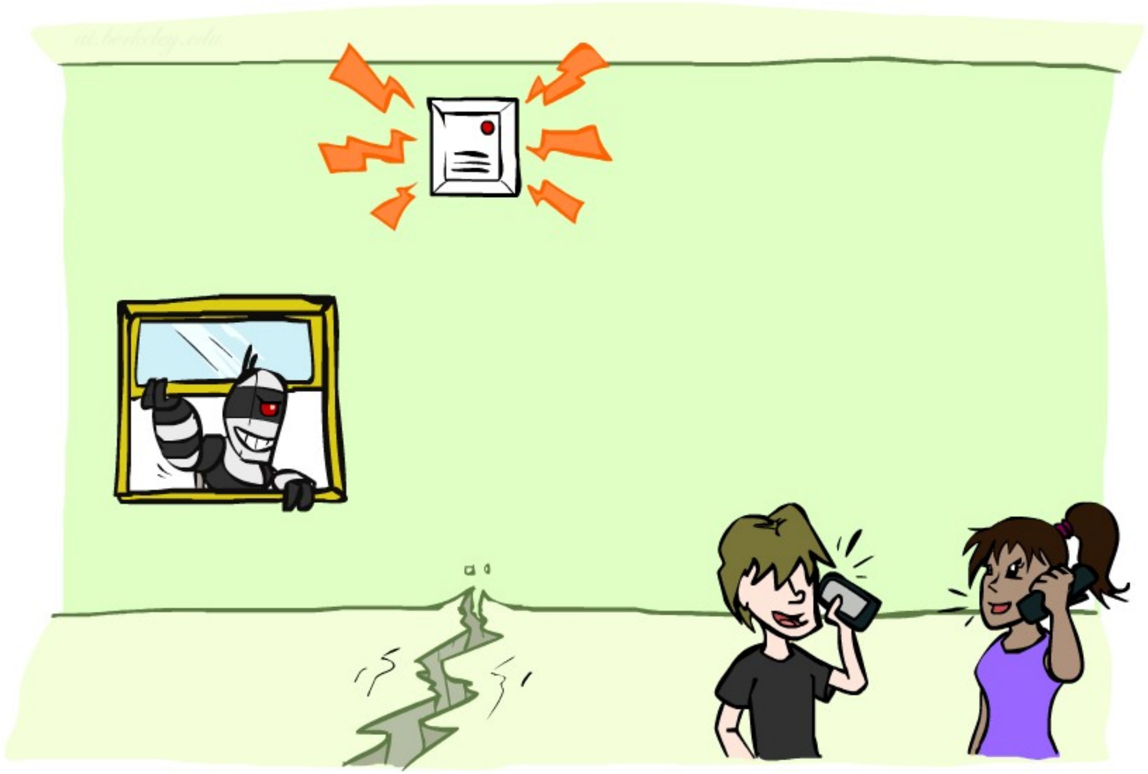
+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

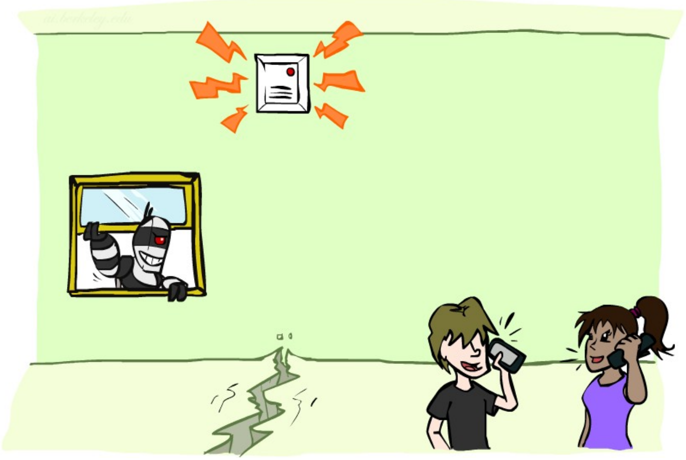
示例 3: Alarm Network

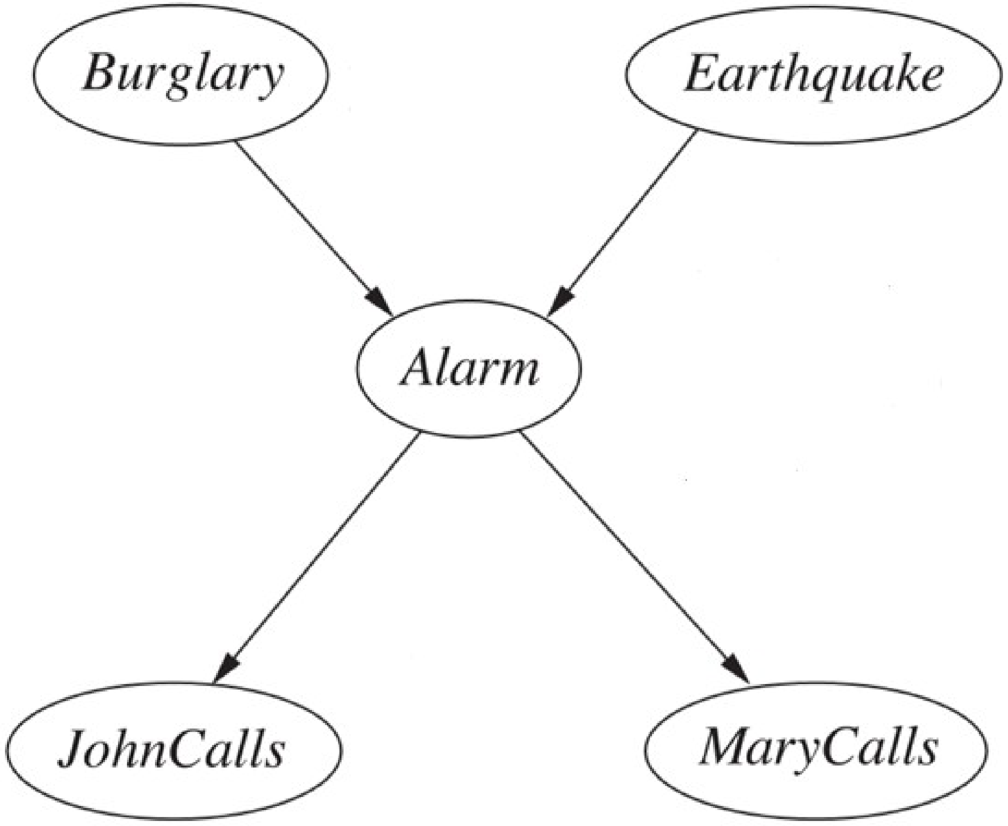
- Alarm Network
  - 我在上班，邻居约翰打电话说我的报警器响了，但邻居玛丽不打电话。有时，它会受到轻微地震的影响。有窃贼吗？
  - 布尔变量：*Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*



示例 3: Alarm Network

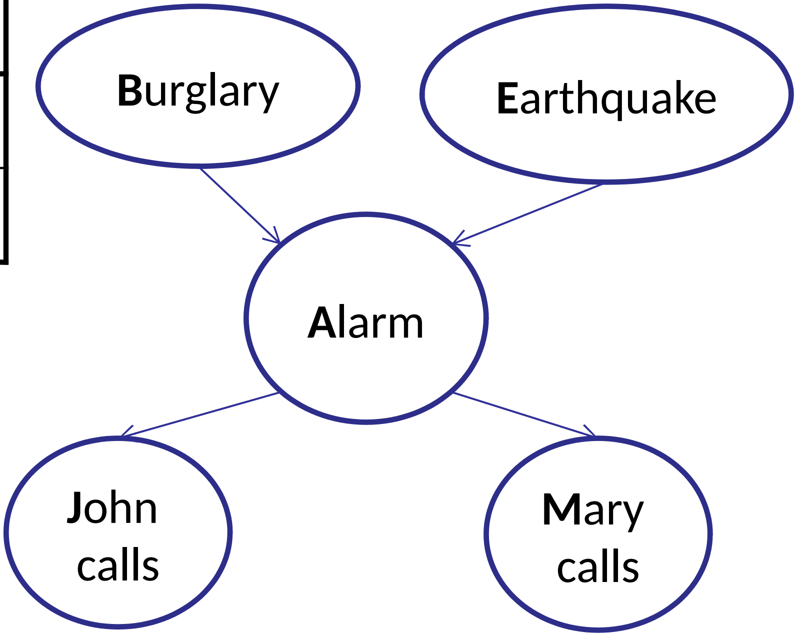
- 布尔变量 : *Burglary, Earthquake, Alarm, JohnCalls, MaryCalls*
- 网络拓扑反应了因果关系
  - 窃贼闯入，报警器会响 *Burglary -> Alarm*
  - 地震偶尔报警器也会有反应 *Earthquake -> Alarm*
  - 听到警报声， John 会打来电话 *Alarm -> JohnCalls*
  - 听到警报声， Mary 会打来电话 *Alarm -> MaryCalls*



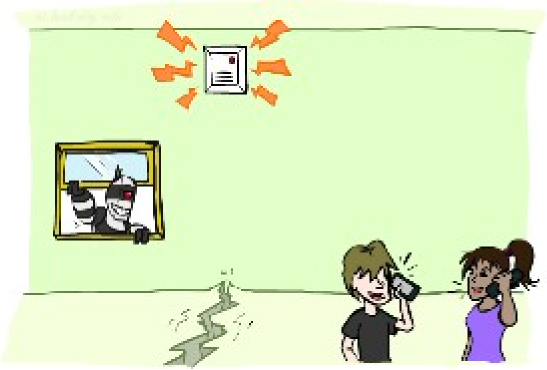


示例 3: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

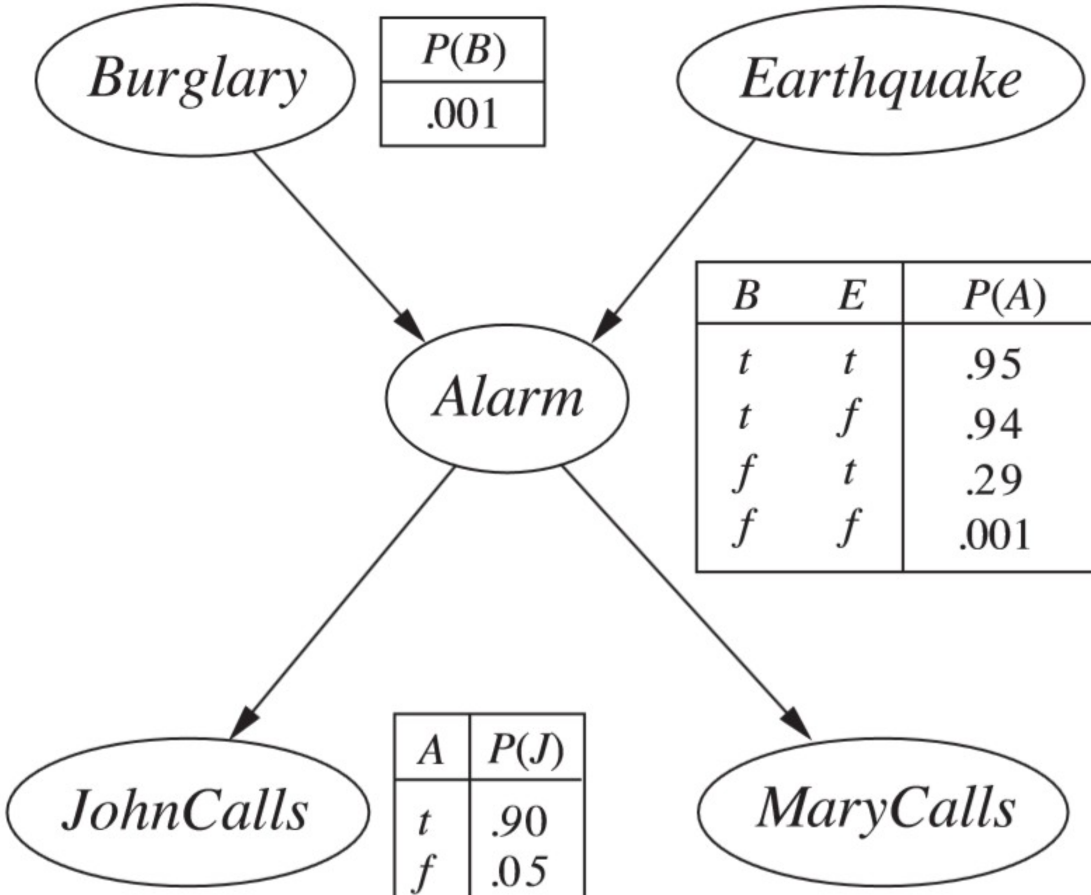


A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

示例 3: Alarm Network



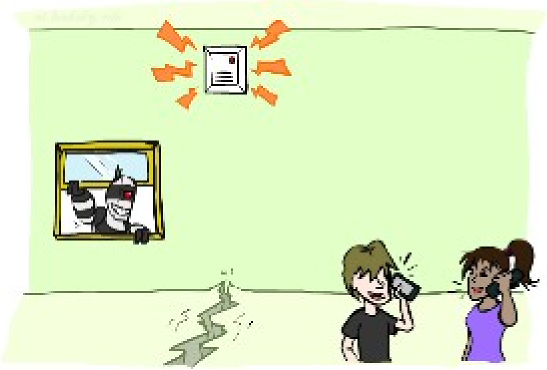
$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

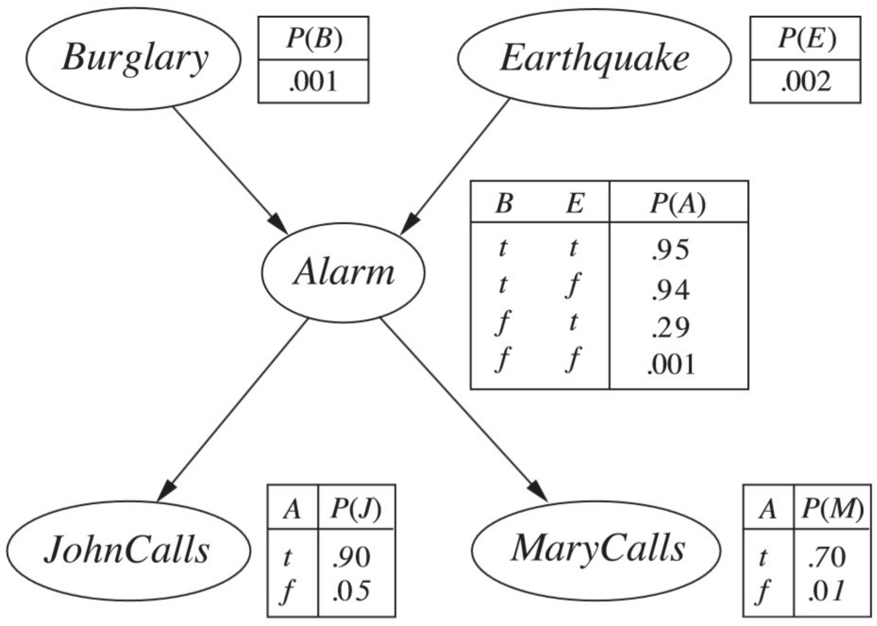


提纲

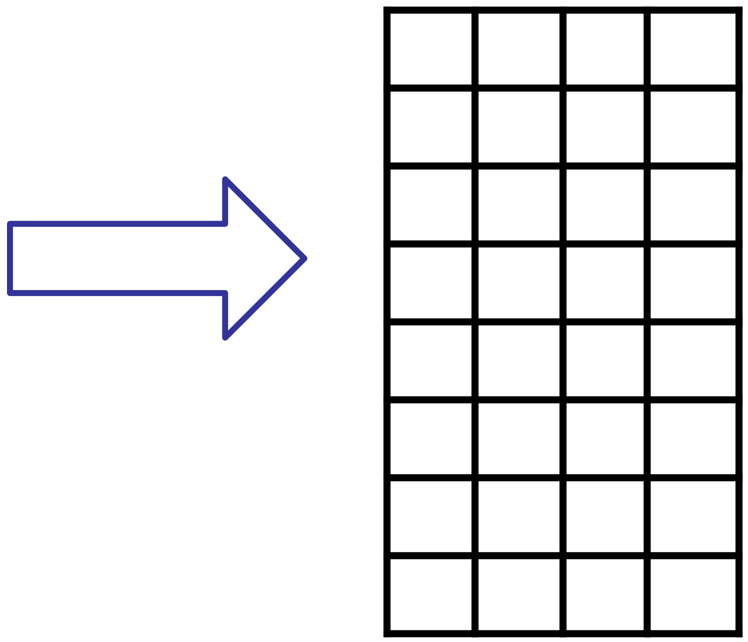
- 第十四章 概率推理
  - 14.1 独立性与条件独立性
  - 14.2 不确定问题的知识表示 - 贝叶斯网络
  - 14.3 贝叶斯网络的语义
  - 14.4 精确推理：枚举推理、 变量消元

贝叶斯网络的语义

- 贝叶斯网络是完全联合概率分布的一种表示



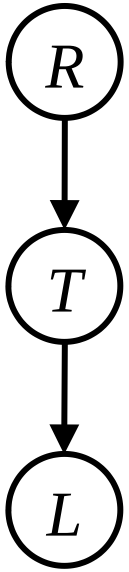
贝叶斯网络



完全联合概率分布

示例：Multiple Joins

$P(R,T,L)=P(R,T)\ P(L|R,T)$



$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

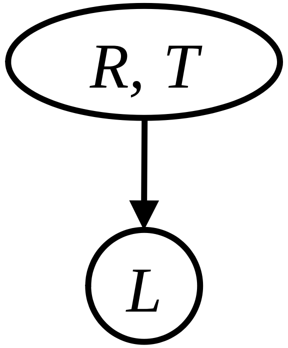
Join R

$P(R,T)$

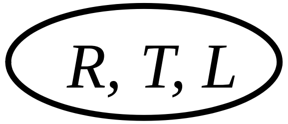
+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9



Join T



$P(R,T,L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

贝叶斯网络的语义

- 贝叶斯网络是完全联合概率分布的一种表示
  - 完全联合概率分布由局部条件概率分布的乘积进行定义（链式法则 + 条件独立性）

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i))$$

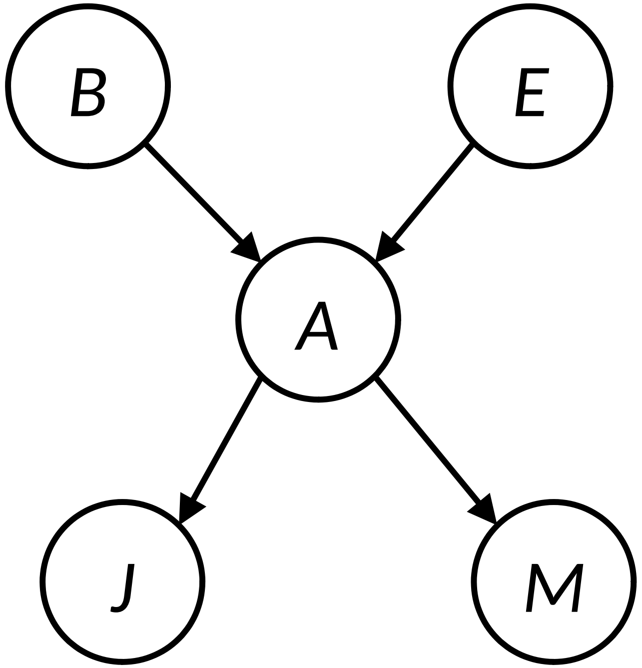
X_i 是 x_i 对应的随机变量

链式法则：

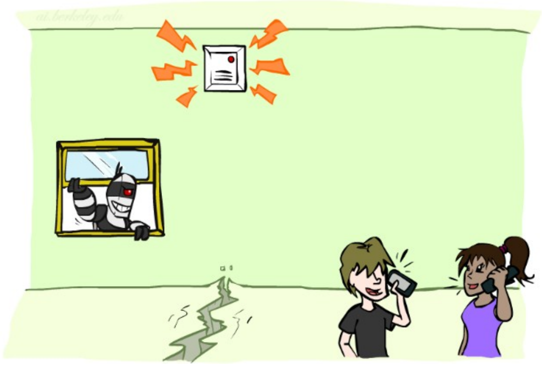
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

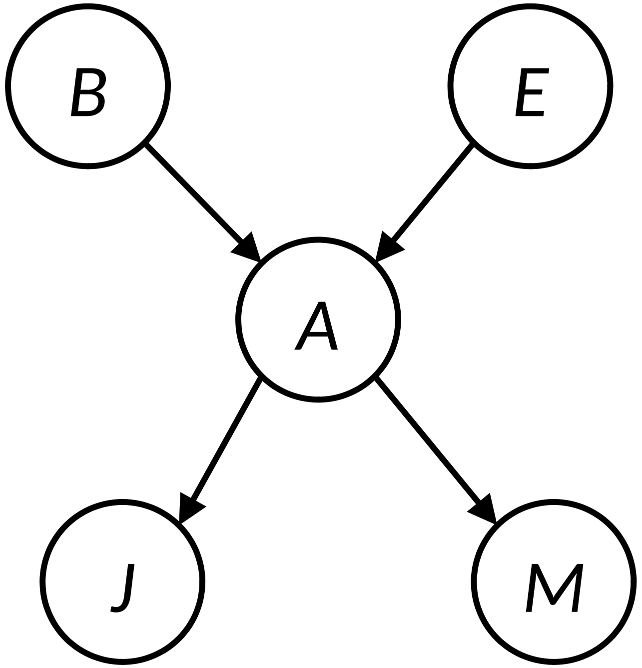
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

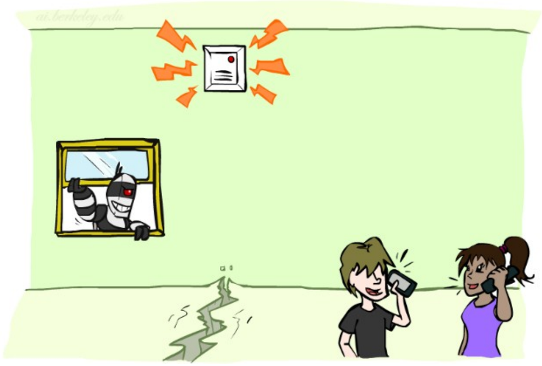
$$P(+b, -e, +a, -j, +m) =$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

$$P(+b, -e, +a, -j, +m) =$$

$$P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) =$$

$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

构造贝叶斯网络的方法

- 1. 结点：选择一组排好序的随机变量 $X_1,\ldots,X_n$
- 2. 边：i 从 1 到 n，执行：

  2.1 从 $X_1,\ldots,X_{i-1}$ 中选择 X_i 的父结点最小的集合，使得

  $$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$$

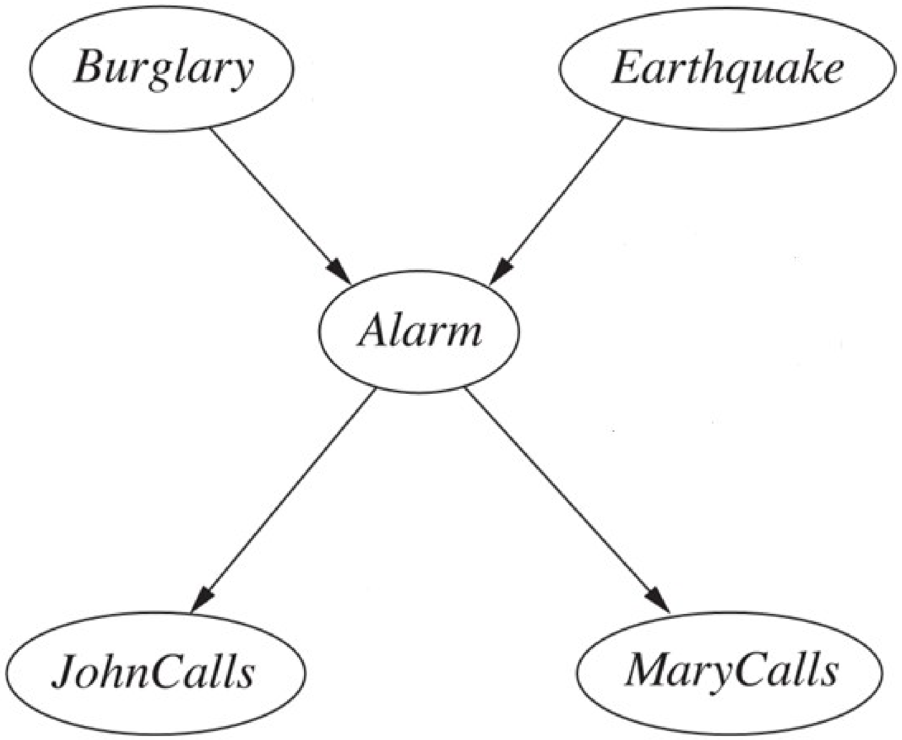
  2.2 在每一个父结点与 X_i 之间插入一条边

  2.3 条件概率表（CPTs）：写出条件概率表 $\mathbf{P}(X_i \mid Parents(X_i))$

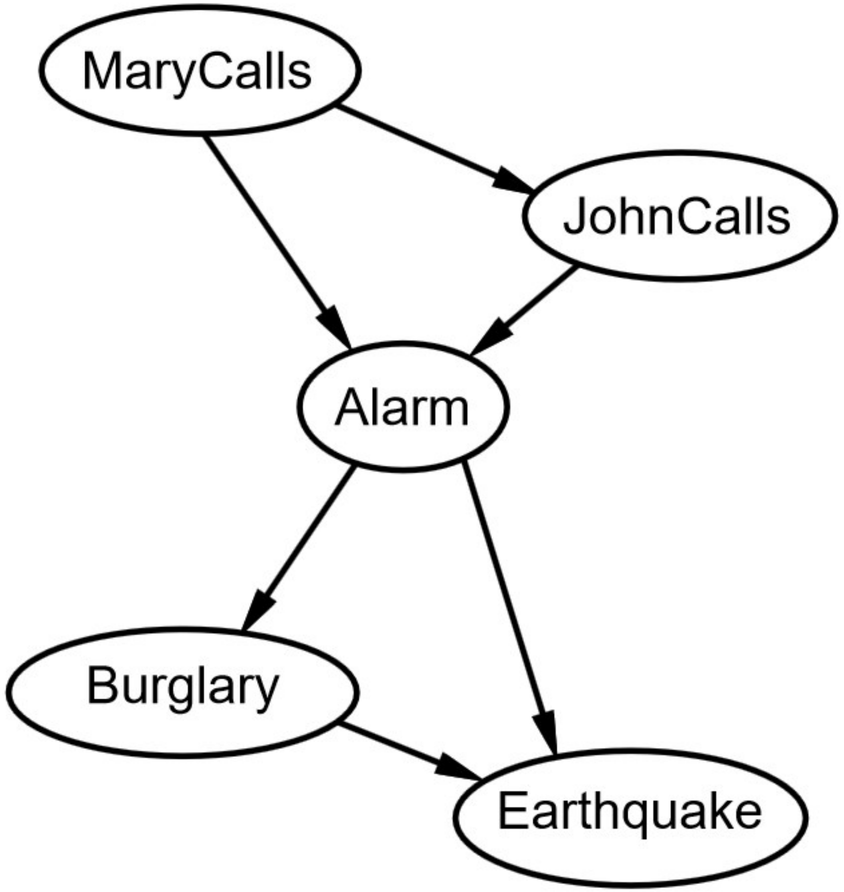
构造贝叶斯网络的方法

示例： Alarm Network

假设选定随机变量的次序： M, J, A, B, E



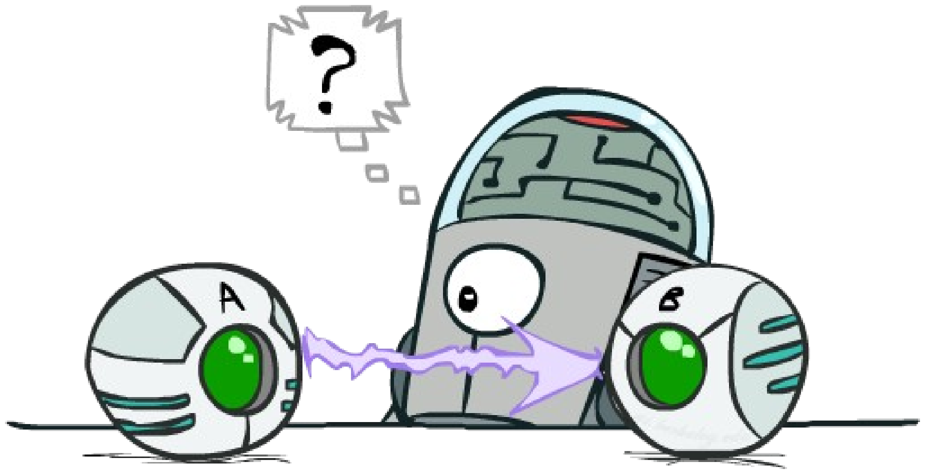
B, E, A, J, M



Causality?

- 当贝叶斯网反映出真正的因果模式时：
  - 通常更简单（节点的父节点更少）
  - 通常更容易思考
  - 通常更容易从专家那里构造出 BNs
- 边的箭头反映相关性，而不是因果关系
  - 拓扑可以编码因果结构
  - 拓扑结构真实反映的是随机变量之间的条件独立性

$$P(x_i|x_1,\ldots x_{i-1}) = P(x_i|parents(X_i))$$



联合分布 VS. Bayes Net

- N 个布尔变量的联合分布有多大？

  2^N

- 如果节点最多有 k 个父节点，那么 N 节点网络有多大？

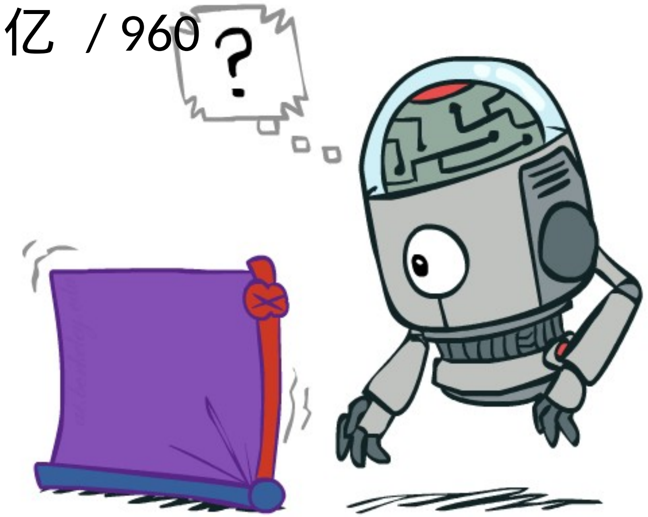
  $O(N * 2^k)$

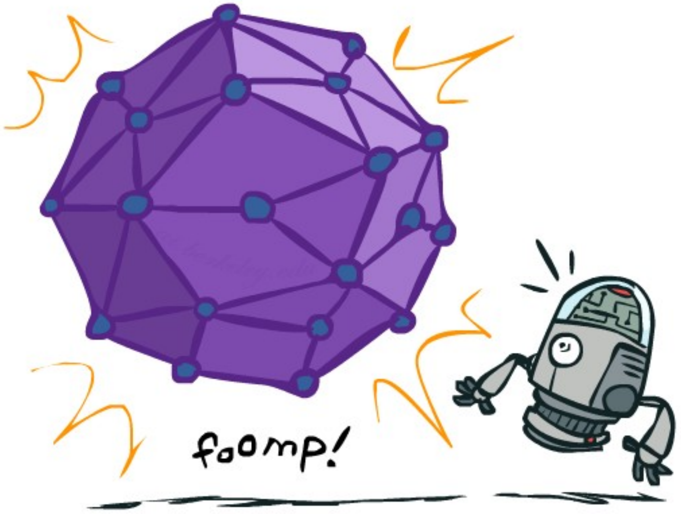
设 N=30 ， k=5 ， 10 亿 / 960

- 两者相同计算的能力

$$P(X_1, X_2, \ldots X_n)$$

- BNs ：节省大量空间！
- 本地 CPT 容易构造





提纲

- 第十四章 概率推理
  - 14.1 独立性与条件独立性
  - 14.2 不确定问题的知识表示 - 贝叶斯网络
  - 14.3 贝叶斯网络的语义
  - 14.4 精确推理：枚举推理、 变量消元

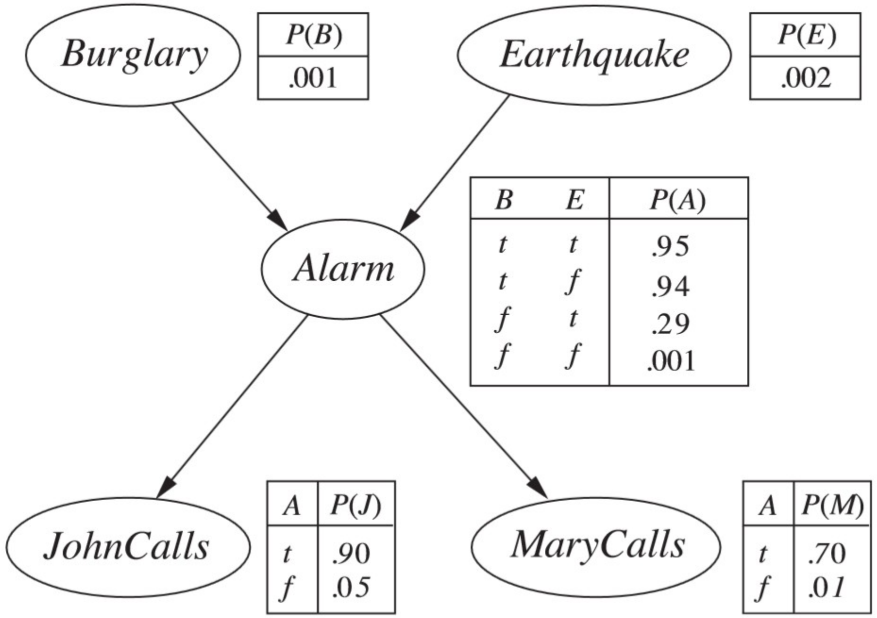
贝叶斯推理

- Inference:
  - 从联合概率分布计算一些有用的量

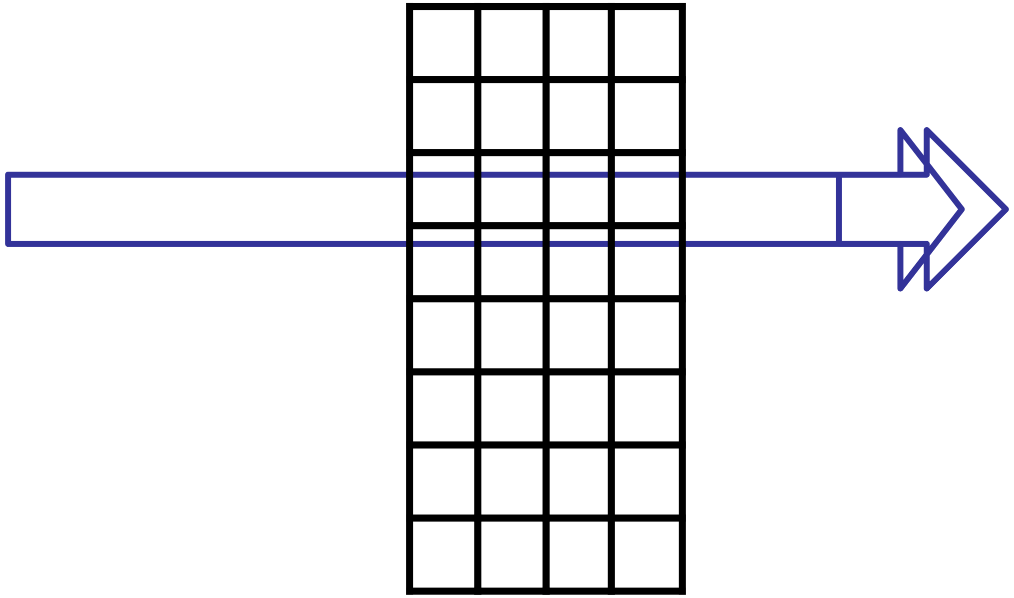
- Examples:
  - 后验概率

$$P(Q|E_1 = e_1, \dots E_k = e_k)$$

贝叶斯网络



联合分布



查询

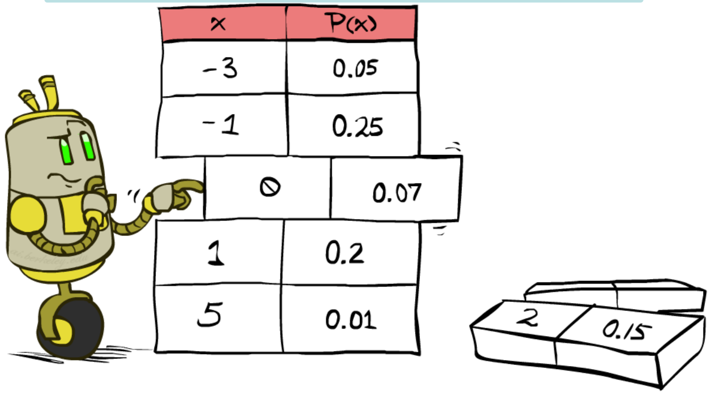
$$P(a \mid e)$$

贝叶斯推理

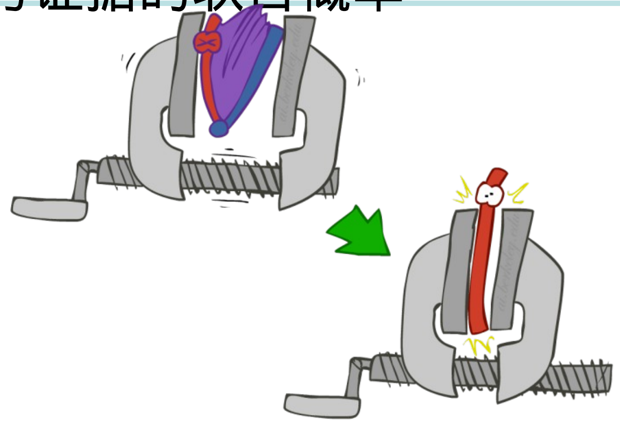
- 基本任务：
  - 证据变量：$E_1 \dots E_k = e_1 \dots e_k$
  - 查询变量：Q
  - 隐藏变量：$H_1 \dots H_r$

$$\left.\begin{matrix} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{matrix}\right\} \begin{matrix} X_1, X_2, \dots X_n \\ \textit{All variables} \end{matrix}$$

- Step 1: 选择与证据一致的条目



- Step 2: 对 H 求和，得到查询与证据的联合概率



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- 典型的查询是询问后验概率：

$$P(Q|e_1 \dots e_k)$$

- Step 3: 归一化

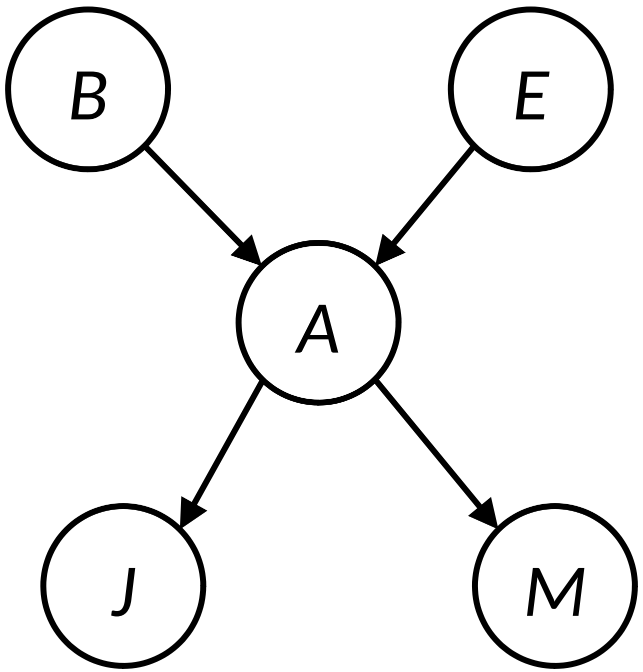
$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

贝叶斯推理

- 示例： Alarm Network
- 已知John和Mary都打来了电话，问出现盗贼的概率分布？
- 分析：
  - 证据变量: *JohnCalls* =true，*MaryCalls* =true
  - 查询变量: Burglary
  - 隐藏变量: *Earthquake*, *Alarm*
- 问题：$P(Burglary|JohnCalls = \text{true}, MaryCalls = \text{true}) = ?$
- 简写为：$P(B|j,m)$?



通过枚举进行推理

- 直接求和联合概率分布中的变量，而不进行实际概率的显式计算

$$\mathbf{P}(B|j,m)$$

根据归一化方法

$$=\alpha\,\mathbf{P}(B,j,m)$$

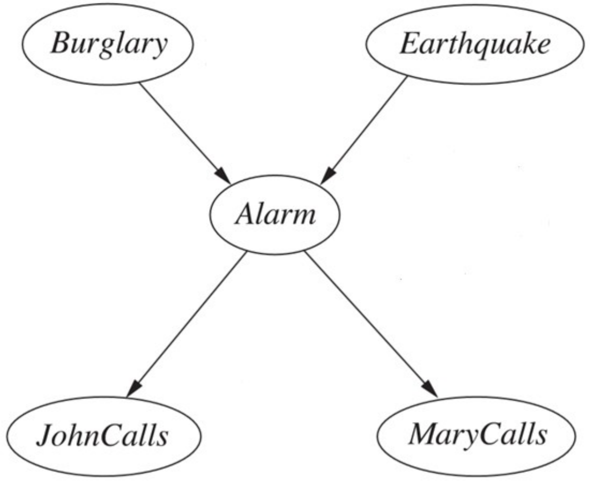
可能世界的概率之和

（隐藏变量求和消元）

$$=\alpha\sum_{e}\sum_{a}\mathbf{P}(B,e,a,j,m)$$

通过枚举进行推理

- 直接求和联合概率分布中的变量，而不进行实际概率的显式计算



$$\mathbf{P}(B|j,m)$$

根据归一化方法

$$=\alpha\,\mathbf{P}(B,j,m)$$

可能世界的概率之和

$$=\alpha\sum_{e}\sum_{a}\mathbf{P}(B,e,a,j,m)$$

根据贝叶斯网络的语义

$$=\alpha\sum_{e}\sum_{a}\mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)$$

简化计算

$$=\alpha\,\mathbf{P}(B)\sum_{e}P(e)\sum_{a}\mathbf{P}(a|B,e)P(j|a)P(m|a)$$

通过枚举进行推理

示例： Alarm Network

$\mathbf{P}(B|j,m)$

$= \alpha\, \mathbf{P}(B) \sum_{e} P(e) \sum_{a} P(j|a)\, P(m|a)\, \mathbf{P}(a|B,e)$

$= \alpha\, \mathbf{P}(B)\ \{P(+e)\{P(+j|+a)P(+m|+a)\mathbf{P}(+a|B,+e) + P(+j|-a)P(+m|-a)\mathbf{P}(-a|B,+e)\}$

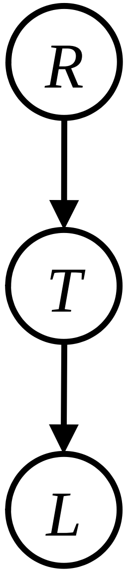
$+ P(-e)\{P(+j|+a)P(+m|+a)\mathbf{P}(+a|B,-e) + P(+j|-a)P(+m|-a)\mathbf{P}(-a|B,-e)\}\}$

$= \alpha < 0.0006, 0.0015 > \approx <0.284,0.716>$

故，在两个邻居都打来电话的条件下，出现盗贼的概率分布约 <0.284,0.716>

Example: Traffic Domain

- Random Variables
  - R: Raining
  - T: Traffic
  - L: Late for class!



$$P(L) = ?$$

$$= \sum_{r,t} P(r, t, L)$$

$$= \sum_{r,t} P(r)P(t|r)P(L|t)$$

$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

小结

- 由于环境可能是部分可观察的或不确定的，智能体需要处理不确定性。
- 概率是描述不确定知识的一种严格形式。通过条件概率，将命题与智能体自身的知识联系起来
- 给定一个完全联合分布可以计算该问题域中任何命题的概率，但对复杂领域，需要找到一种方法来降低联合概率的数目
- 独立性和条件独立性提供了重要工具
- 贝叶斯网络 = 拓扑结构（图）+ 条件概率表 (CPT)
- 贝叶斯网络是一种使用简单局部分布（条件概率）描述复杂联合分布的技术
- 贝叶斯网络中通过枚举、变量消元算法进行概率推理

谢谢！