



高级机器学习

概率图模型



概率基础

- Conditional probability $P(x|y) = \frac{P(x, y)}{P(y)}$
- Product rule $P(x, y) = P(x|y)P(y)$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x, y) = P(x)P(y)$

- X and Y are conditionally independent given Z if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \quad X \perp\!\!\!\perp Y|Z$$

概率模型

- 机器学习最重要的任务是根据**已观察**到的证据（例如训练样本）对感兴趣的**未知**变量（例如类别标记）进行估计和推测
 - **概率模型**（probabilistic model）提供了一种描述框架，将描述任务归结为**计算变量的概率分布**，在概率模型中，利用**已知**的变量推测**未知**变量的分布称为“推断（inference）”，其**核心**在于基于可观测的变量推测出未知变量的**条件分布**
 - 生成式：计算联合分布 $P(Y, R, O)$
 - 判别式：计算条件分布 $P(Y, R|O)$
-

概率图模型

直接利用概率求和规则消去变量 R 的时间和空间复杂度为指数级别 $O(2^{|Y|+|R|})$ ，需要一种能够简洁紧凑表达变量间关系的工具

概率图模型

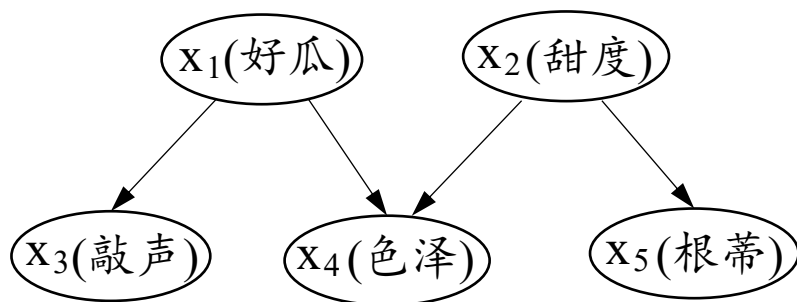
概率模型

- 概率图模型(probabilistic graphical model)是一类用图来表达**变量相关关系**的概率模型
 - 图模型提供了一种**描述框架**,
 - 结点：随机变量（集合）
 - 边：变量之间的依赖关系
 - 分类：
 - **有向图**：贝叶斯网
 - 使用有向无环图表示变量之间的依赖关系
 - **无向图**：马尔可夫网
 - 使用无向图表示变量间的相关关系
-

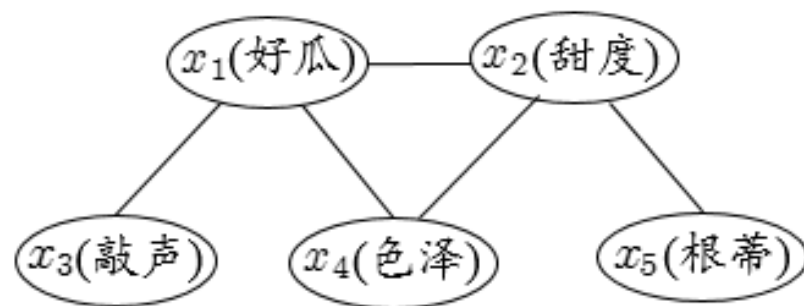
概率模型

□ 概率图模型分类：

- 有向图：贝叶斯网
- 无向图：马尔可夫网



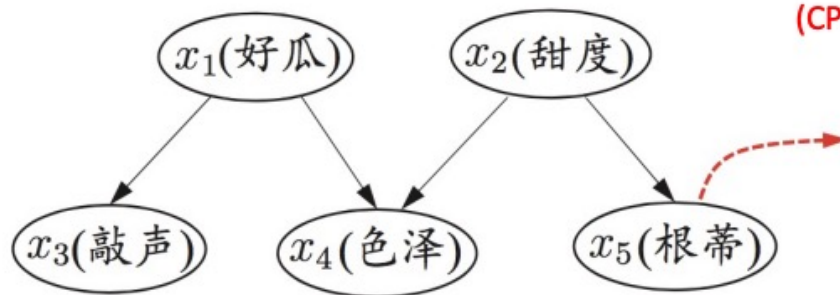
有向图



无向图

贝叶斯网

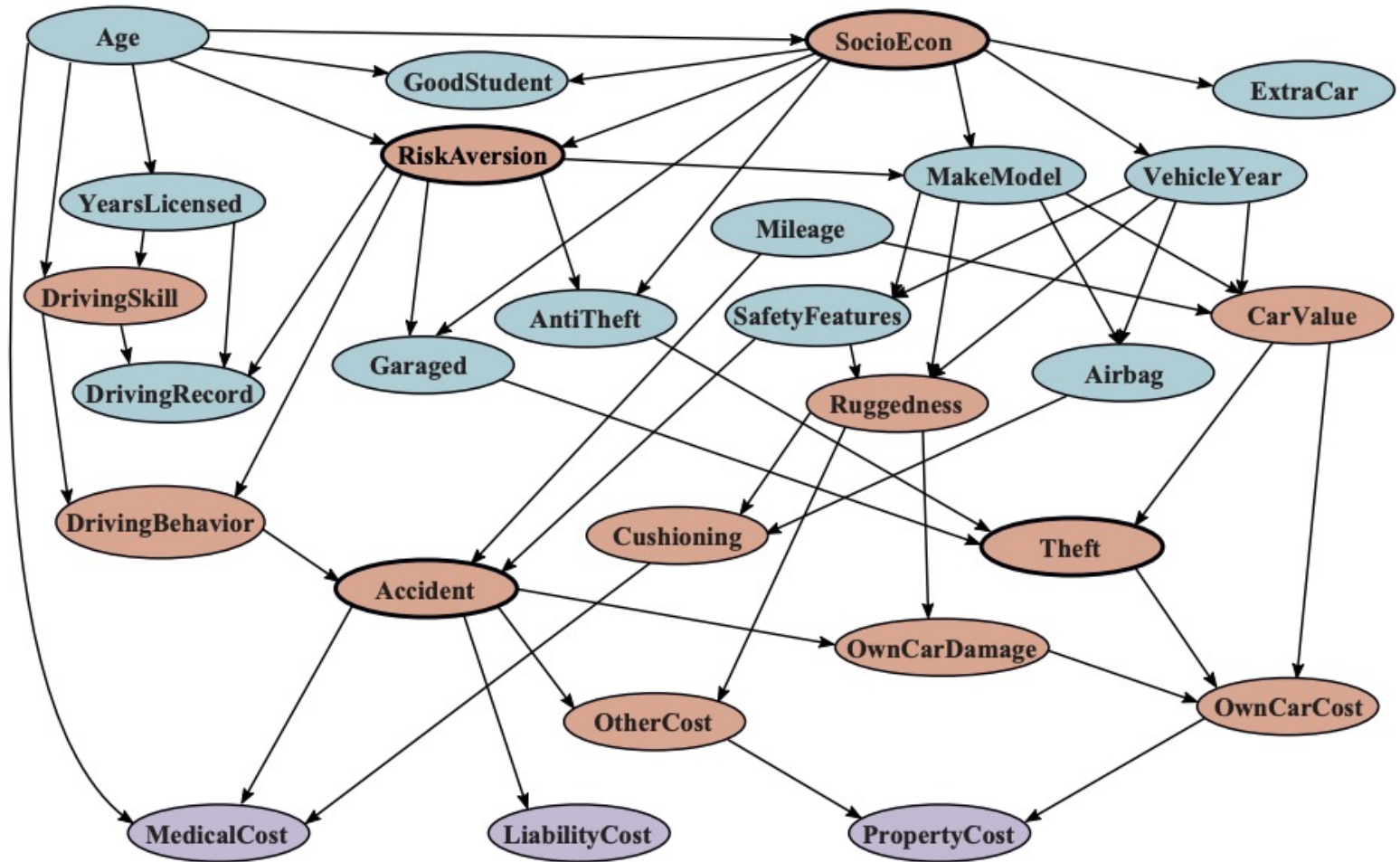
- 每个节点对应一个随机变量
- 一组有向边，没有有向回路（有向无环图，DAG）
- 每个节点有一个条件概率分布 $P(X_i | Parents(X_i))$ ，量化其父节点对该节点的影响
 - 条件概率表 (Conditional Probability Table, CPT)



条件概率表
(CPT, Conditional Probability Table)

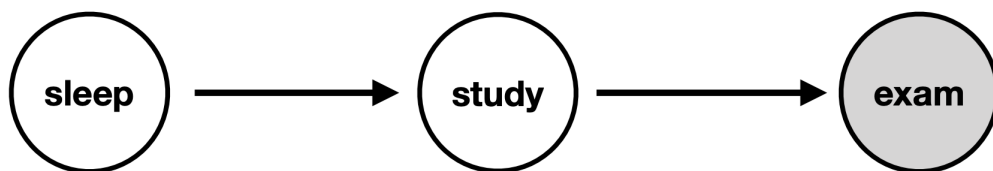
		根蒂	
		硬挺	蜷缩
甜度	高	0.1	0.9
	低	0.7	0.3

贝叶斯网



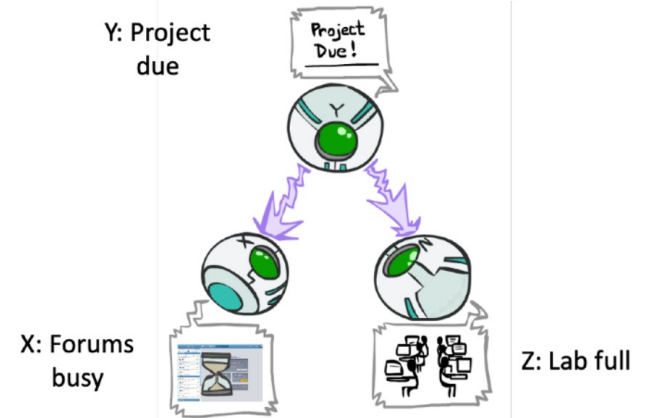
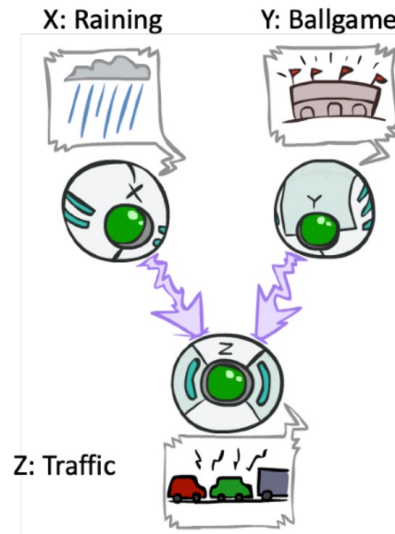
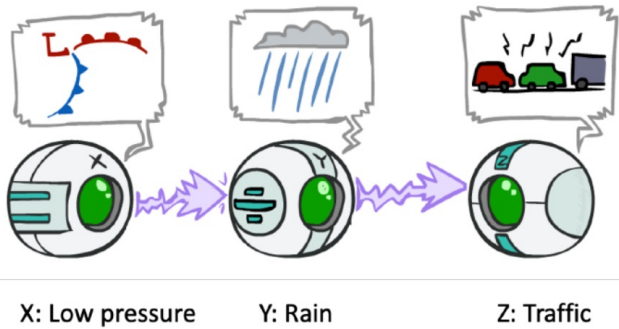
贝叶斯网中的条件独立性

- 贝叶斯网络的重要问题：
 - 两个结点（变量）是否（条件）独立？
- 例如：

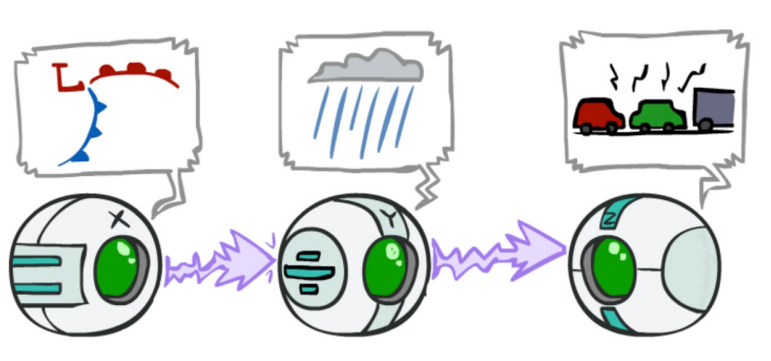


- sleep和exam是否独立？
 - 如何让它们独立？
-

三个变量之间的典型依赖关系



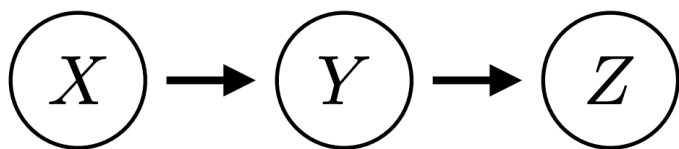
三个变量之间的典型依赖关系



X: Low pressure

Y: Rain

Z: Traffic



$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y)$$

- X和Z是否独立？

NO!

- 直观解释：

- Low pressure导致rain, rain导致traffic
- High pressure 导致 no rain, no rain 导致 no traffic

$$P(+y|+x) = 1, P(-y|-x) = 1$$

$$P(+z|+y) = 1, P(-z|-y) = 1$$

$$P(X, Z) = P(X) \sum_Y P(Y|X)P(Z|Y) = P(X)P(Z|X)$$

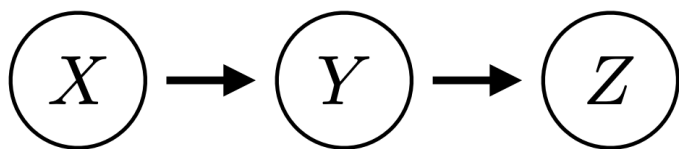
三个变量之间的典型依赖关系



X: Low pressure

Y: Rain

Z: Traffic



$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y)$$

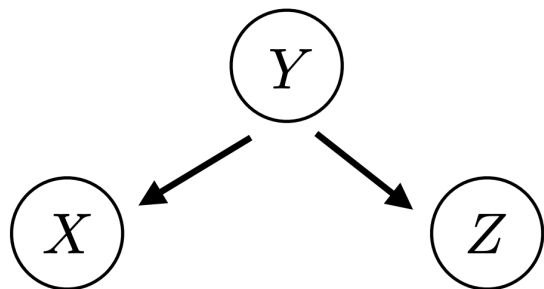
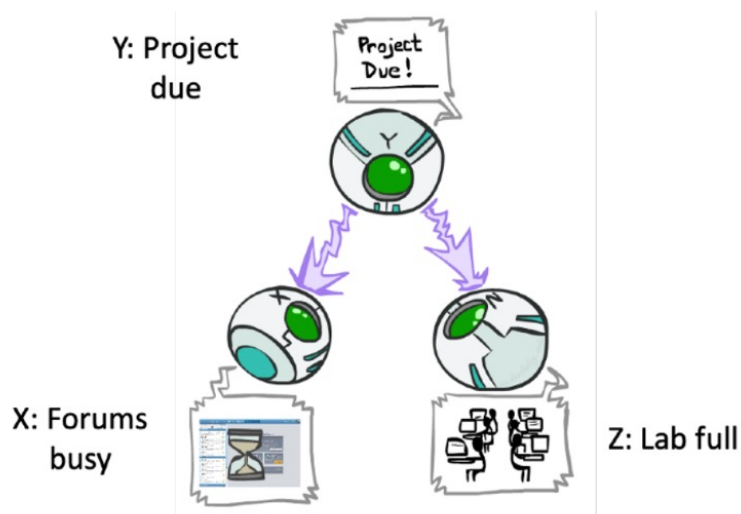
- 给定Y的取值，X和Z是否独立？

$$\begin{aligned} P(X, Z|Y) &= \frac{P(X, Y, Z)}{P(Y)} = \frac{P(X)P(Y|X)P(Z|Y)}{P(Y)} \\ &= \frac{P(Y)P(X|Y)P(Z|Y)}{P(Y)} \\ &= P(X|Y)P(Z|Y) \end{aligned}$$

Yes!

- 观测变量(evidence)阻断(blocks)了影响的传播

三个变量之间的典型依赖关系



- X和Z是否独立？

NO!

- 直观解释：

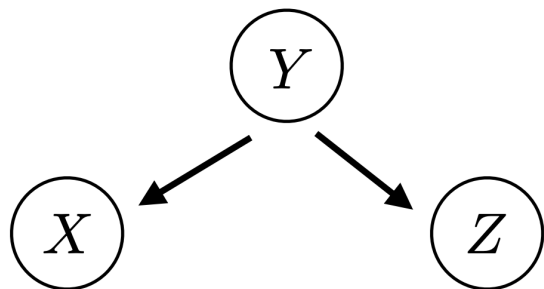
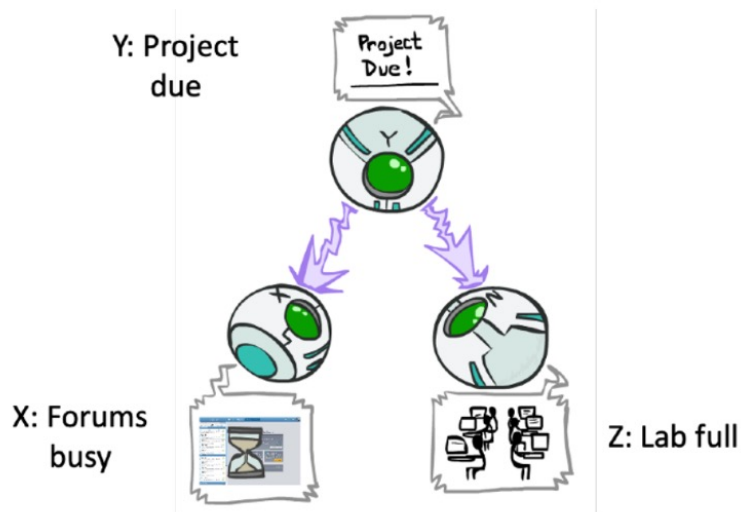
- Project due同时导致 forums busy和lab full

$$P(+x|+y) = 1, P(-x|-y) = 1$$

$$P(+z|+y) = 1, P(-z|-y) = 1$$

$$P(X, Y, Z) = P(Y)P(X|Y)P(Z|Y) \quad P(X, Z) = \sum_Y P(Y)P(X|Y)P(Z|Y) \neq P(X)P(Z)$$

三个变量之间的典型依赖关系



$$P(X, Y, Z) = P(Y)P(X|Y)P(Z|Y)$$

- 给定Y的取值，X和Z是否独立？

$$P(X, Z|Y) = \frac{P(X, Y, Z)}{P(Y)}$$

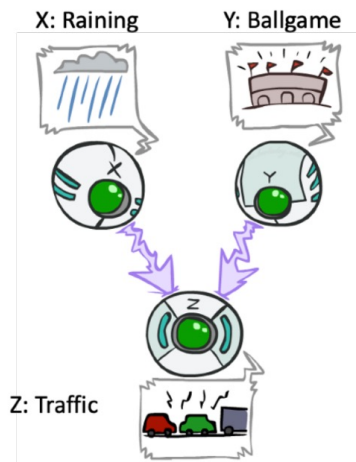
$$= \frac{P(Y)P(X|Y)P(Z|Y)}{P(Y)}$$

$$= P(X|Y)P(Z|Y)$$

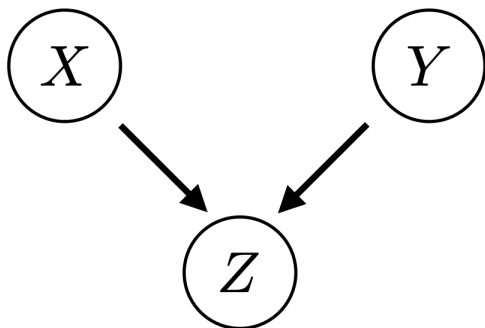
Yes!

- 观察到原因，阻断了影响的传播

三个变量之间的典型依赖关系



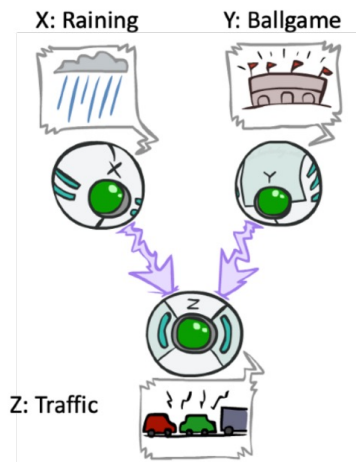
- X和Y是否独立？
- **Yes:** ballgame和raining都导致traffic，但它们本身并不相关



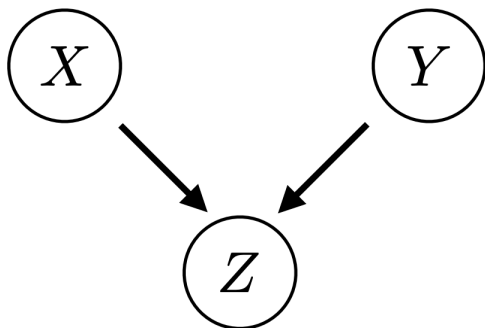
$$P(X, Y) = \sum_Z P(X)P(Y)P(Z|X, Y) = P(X)P(Y)$$

$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$$

三个变量之间的典型依赖关系



- 给定Z, X和Y是否独立？
- No: 观察到traffic, 要么ballgame, 要么raining



$$P(X, Y|Z) = \frac{P(X, Y, Z)}{P(Z)} = \frac{P(X)P(Y)P(Z|X, Y)}{P(Z)}$$

$$\neq P(X|Z)P(Y|Z)$$

$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$$

- 观察到结果, 激活了影响的传播

一般情况

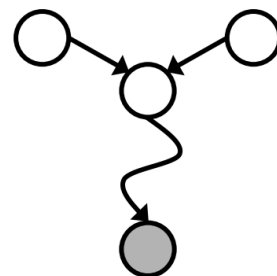
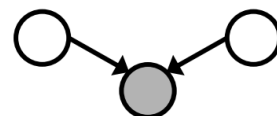
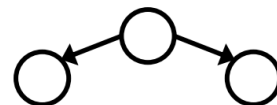
□ 给定观测变量Z，X和Y是否条件独立？

- Yes, 如果X和Y被Z阻断
- 考虑所有X到Y的路径
- 没有联通的边=独立性！

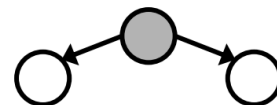
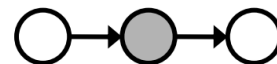
□ 一个边是联通的，如果每个三元组active

- $A \rightarrow B \rightarrow C$ ，如果B未观测，active
- $A \leftarrow B \rightarrow C$ ，如果B未被观测，active
- $A \rightarrow B \leftarrow C$ ，如果B或者B的子节点被观测，active

Active Triples



Inactive Triples



有向分离(D-Separation)

□ 询问: $X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$?

□ 检查 X_i 和 X_j 之间的所有路径

- 如果有一个路径active, 则独立性不成立

$$X_i \not\perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

- 反之, 如果所有路径都是inactive, 则独立性成立

$$X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

有向分离(D-Separation)

□ 询问: $X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$?

□ 检查 X_i 和 X_j 之间的所有路径

- 如果有一个路径active, 则独立性不成立

$$X_i \not\perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

- 反之, 如果所有路径都是inactive, 则独立性成立

$$X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

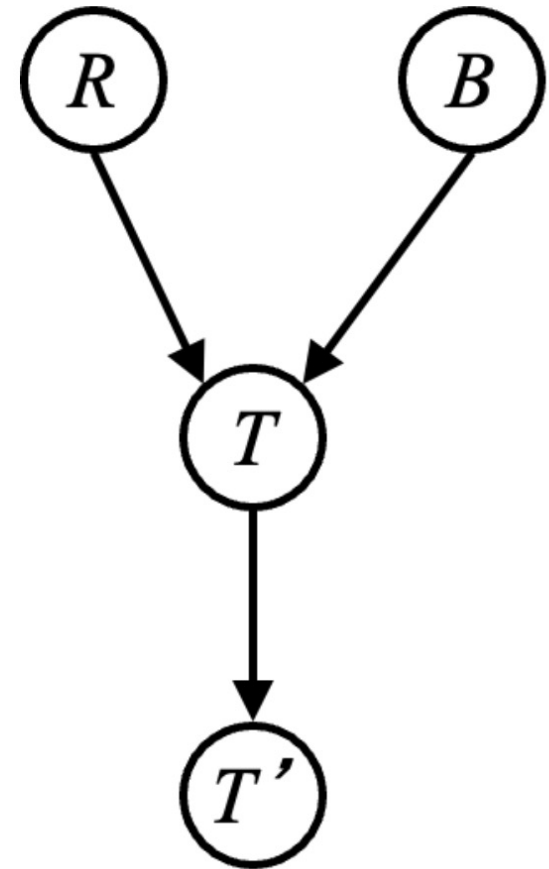
Example

$$R \perp\!\!\!\perp B$$

Yes

$$R \perp\!\!\!\perp B | T$$

$$R \perp\!\!\!\perp B | T'$$



Example

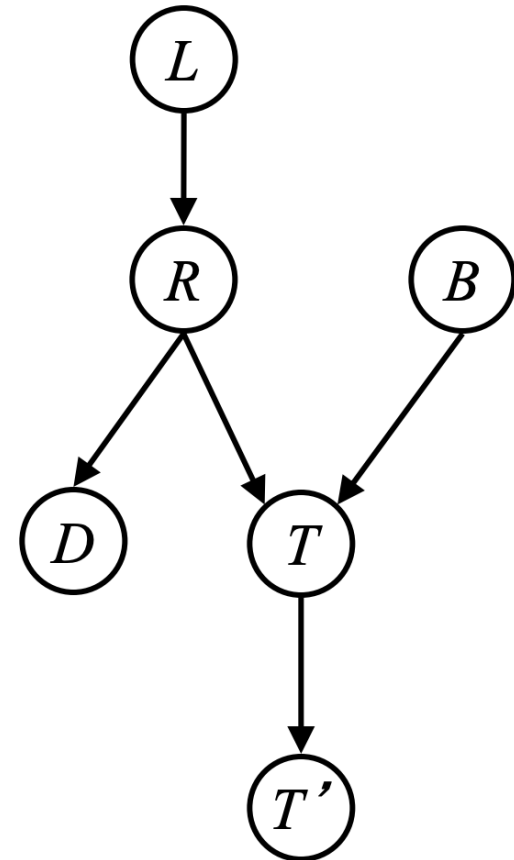
$L \perp\!\!\!\perp T' | T$ Yes

$L \perp\!\!\!\perp B$ Yes

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ Yes

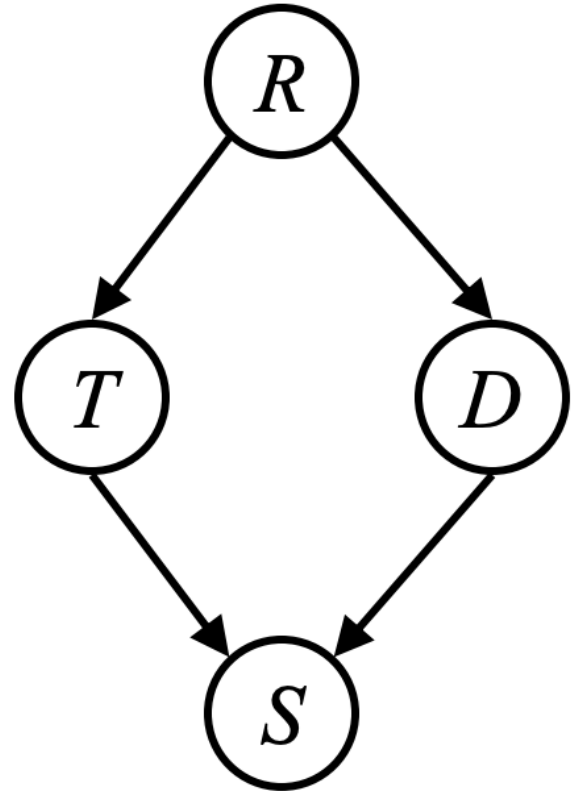


Example

$$T \perp\!\!\!\perp D$$

$$T \perp\!\!\!\perp D \mid R \quad \text{Yes}$$

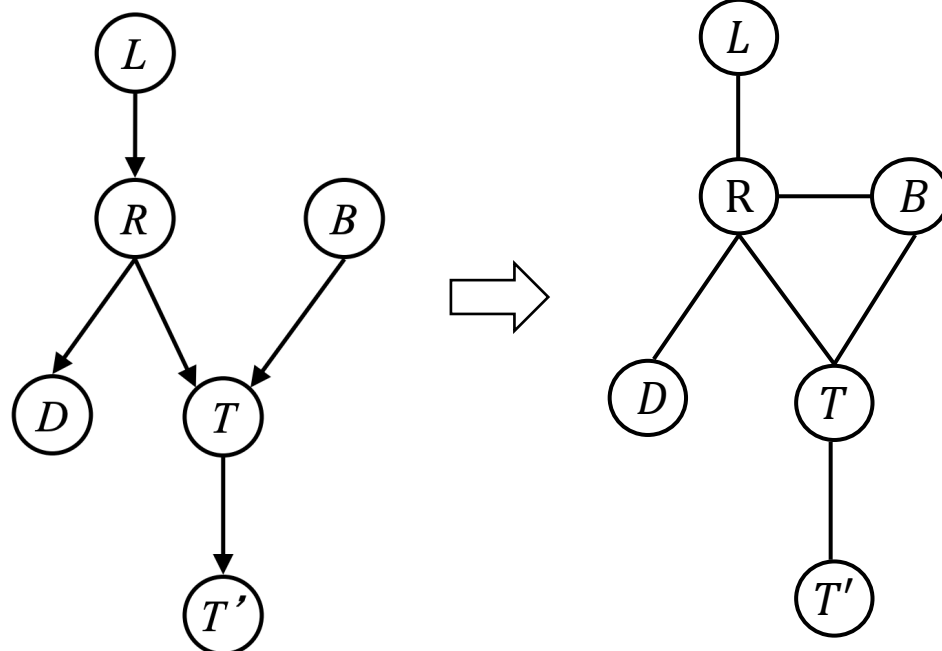
$$T \perp\!\!\!\perp D \mid R, S$$



道德图(moral graph)

□ 先将有向图转变为无向图

- V型结构父结点相连
- 有向边变成无向边



□ 从图中将观测变量 z 去除后，如果 x ， y 属于两个联通分支，则条件独立性成立

$$L \perp\!\!\!\perp T' | T \quad \text{Yes}$$

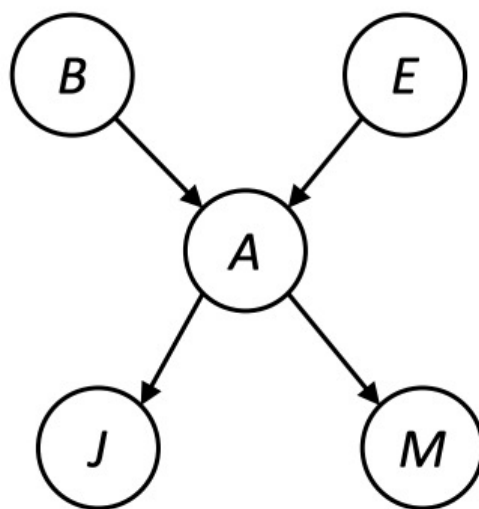
$$L \perp\!\!\!\perp B | T$$

$$L \perp\!\!\!\perp B | T'$$

$$L \perp\!\!\!\perp B | T, R \quad \text{Yes}$$

贝叶斯网

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

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

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



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

$$\begin{aligned}
 P(+b, -e, +a, -j, +m) &= \\
 P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) &= \\
 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7 &=
 \end{aligned}$$

推理任务

- 概率推理系统的基本任务：给定某个已观察到的事件，即一组证据变量(evidence variables)的取值后，计算一组查询变量(query variable)的后验概率分布
- 例如，在防盗贝叶斯网络中，观察到事件：JohnCalls = true 并且 MaryCalls=true，问出现小偷的概率是多少：

$$P(\text{Burglary} | \text{JohnCalls} = \text{true}, \text{MaryCalls} = \text{true}) = ?$$

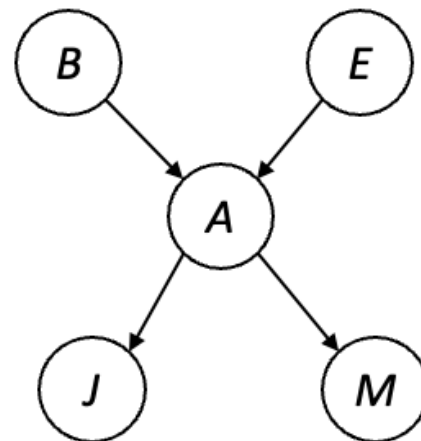
推理任务

$$P(B|j, m) = \alpha P(B, j, m)$$

$$= \alpha \sum_e \sum_a P(B, j, m, e, a)$$

$$= \alpha \sum_e \sum_a P(B)P(e)P(a|B, e)P(j|a)P(m|a)$$

$$\begin{aligned} &= P(B)P(e)P(a|B, e)P(j|a)P(m|a) \\ &+ P(B)P(e)P(\neg a|B, e)P(j|\neg a)P(m|\neg a) \\ &+ P(B)P(\neg e)P(a|B, \neg e)P(j|a)P(m|a) \\ &+ P(B)P(e)P(\neg a|B, e)P(j|\neg a)P(m|\neg a) \end{aligned}$$



变量消元算法

- 为什么枚举推理复杂度这么高？
 - 隐变量消去的太迟了
 - 变量消元(Variable elimination)算法
 - 尽早消去隐变量
-

变量消元算法

- 变量消元算法：按照从右到左的顺序计算，保存中间结果避免重复计算

$$\begin{aligned} P(B|j, m) &= \alpha \sum_{e, a} P(B)P(e)P(a|B, e)P(j|a)P(m|a) \\ &= \underbrace{\alpha P(B)}_{f_1(B)} \underbrace{\sum_e P(e)}_{f_2(E)} \underbrace{\sum_a P(a|B, e)}_{f_3(A, B, E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)} \end{aligned}$$

因子 (factors)

变量消元算法

$$\begin{aligned} & P(B|j, m) \\ &= \alpha \sum_{e, a} P(B)P(e)P(a|B, e)P(j|a)P(m|a) \\ &= \alpha \underbrace{P(B)}_{f_1(B)} \underbrace{\sum_e P(e)}_{f_2(E)} \underbrace{\sum_a P(a|B, e)}_{f_3(A, B, E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)} \\ &= \alpha P(B) \sum_e P(e) \sum_a f_3(A, B, E) f_4(A) f_5(A) \\ &= \alpha P(B) \sum_e P(e) f_6(B, E) \quad (\text{对A进行求和消元}) \\ &= \alpha P(B) \sum_e f_2(E) f_6(B, E) = f_7(B) \quad (\text{对E进行求和消元}) \\ &= \alpha f_1(B) f_7(B) \end{aligned}$$

精确推理

□ 单连通图(singly connected networks)或多形树(polytrees)

- 任意两个结点之间最多只有一条无向路径
- 时间和空间复杂度： $O(d^k n)$

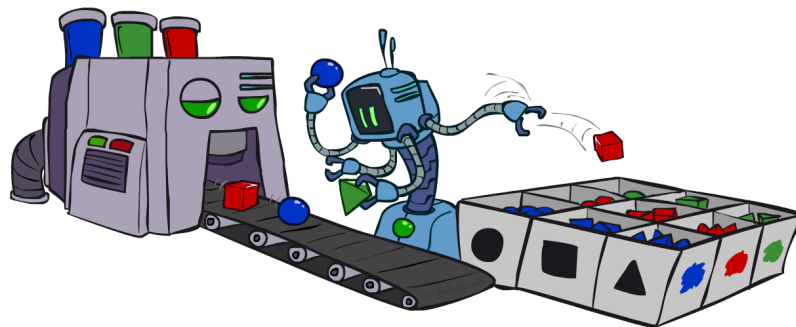
□ 多连通图(multi connected networks)

- NP-Hard
-

采样

□ 基本思想：

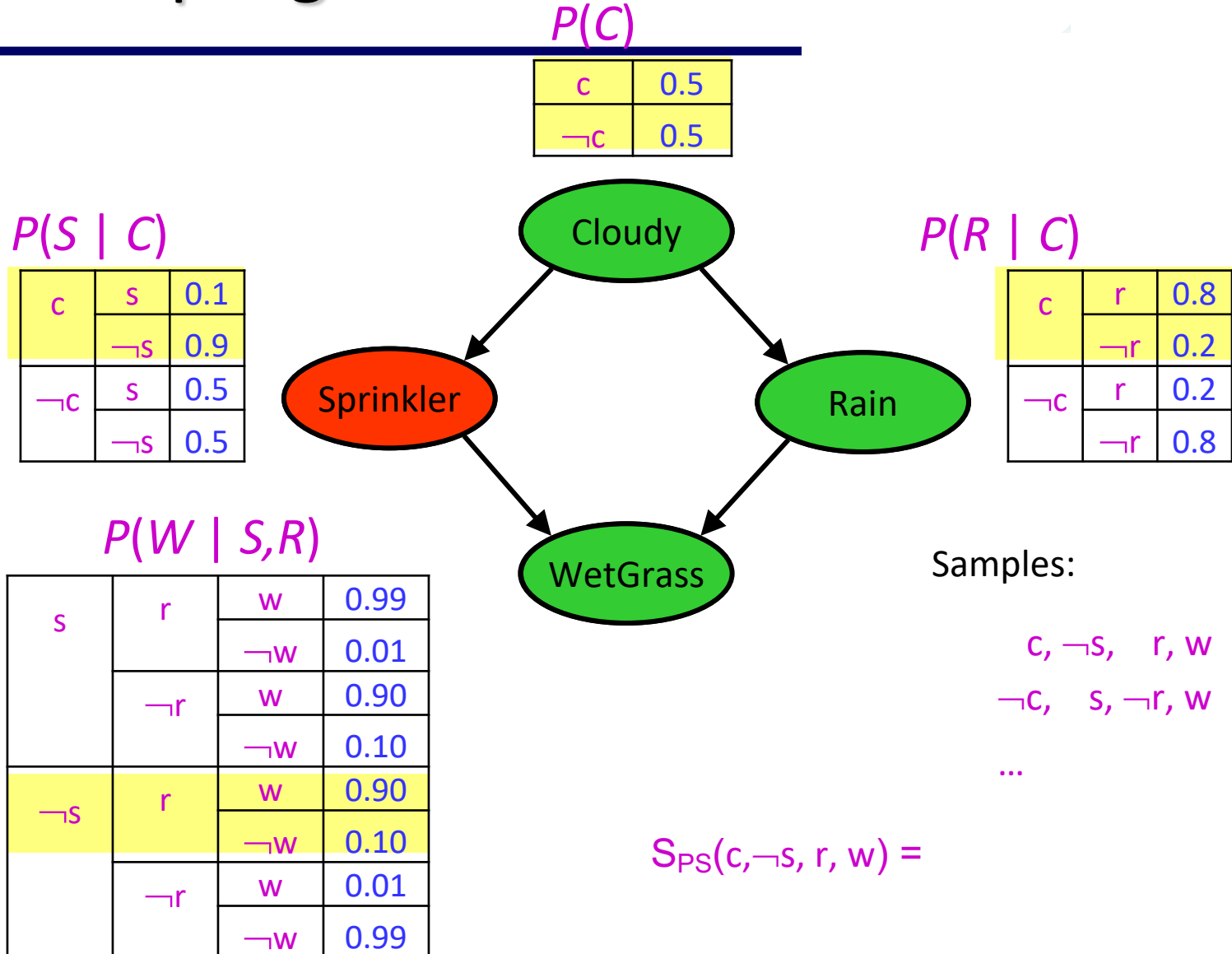
- 从分布 S 采样 N 个样本
- 计算近似的概率 $\hat{P}$
- 说明近似概率收敛于真实概率 P



采样

- 直接采样 (Prior Sampling)
 - 拒绝采样 (Rejection Sampling)
 - 似然加权 (Likelihood Weighting)
 - 吉布斯采样 (Gibbs Sampling)
-

Prior Sampling



Prior Sampling

- 采样算法生成特定事件的概率：

$$S_{PS}(x_1, \dots, x_n) = \prod_i P(x_i | \text{parents}(x_i)) = P(x_1, \dots, x_n)$$

例如： $S_{PS}(c, \neg s, r, w) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(c, \neg s, r, w)$

- 假设采样生成 N 个样本，令 $N_{PS}(x_1, \dots, x_n)$ 为特定事件 $(x_1, \dots, x_n)$ 出现的次数

$$\begin{aligned} \lim_{N \rightarrow \infty} \hat{P}(x_1, \dots, x_n) &= \lim_{N \rightarrow \infty} N_{PS}(x_1, \dots, x_n) / N \\ &= S_{PS}(x_1, \dots, x_n) \\ &= P(x_1 \dots x_n) \end{aligned}$$

采样算法满足一致性(consistent)

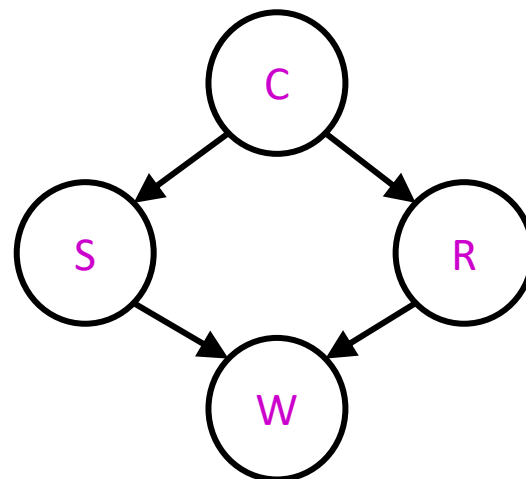
Prior Sampling

□ 从贝叶斯网中采样得到下列样本

$C, \neg S, r, W$
 C, S, r, W
 $\neg C, S, r, \neg W$
 $C, \neg S, r, W$
 $\neg C, \neg S, \neg r, W$

□ $P(W)$?

- 计数: $\langle w:4, \neg w:1 \rangle$
- 归一化: $P(W) = \langle w:0.8, \neg w:0.2 \rangle$



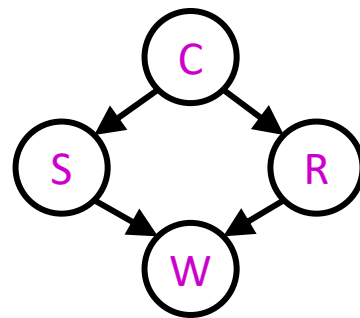
采样的样本越多，估计越准确

□ $P(C, +w)$? $P(C|+r, +w)$? $P(C|\neg r, \neg w)$

拒绝采样 (Rejection Sampling)

□ 用于计算条件概率：

- 计算 $P(C|r, w)$
- $\neg r$ 和 $\neg w$ 不匹配
- 计数时只需要计算匹配的样本，拒绝其它样本

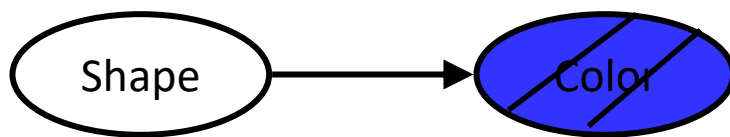


$C, \neg S, r, w$
 ~~$C, S, \neg r$~~
 ~~$\neg C, S, r, \neg w$~~
 ~~$C, \neg S, \neg r$~~
 $\neg C, \neg S, r, w$

问题：拒绝了太多的样本，随着证据变量的增多，与证据相一致的样本数量指数级下降

似然加权(Likelihood weighting)

- 基本思想：只生成与证据一致的事件
 - 问题：采样分布不一致
 - 方法：在对查询样本的分布计数时，每个事件以它与证据吻合的似然（相似性）为权值



pyramid, blue
pyramid, blue
sphere, blue
cube, blue
sphere, blue

似然加权(Likelihood weighting)

$P(C)$

c	0.5
$\neg c$	0.5

$P(S | C)$

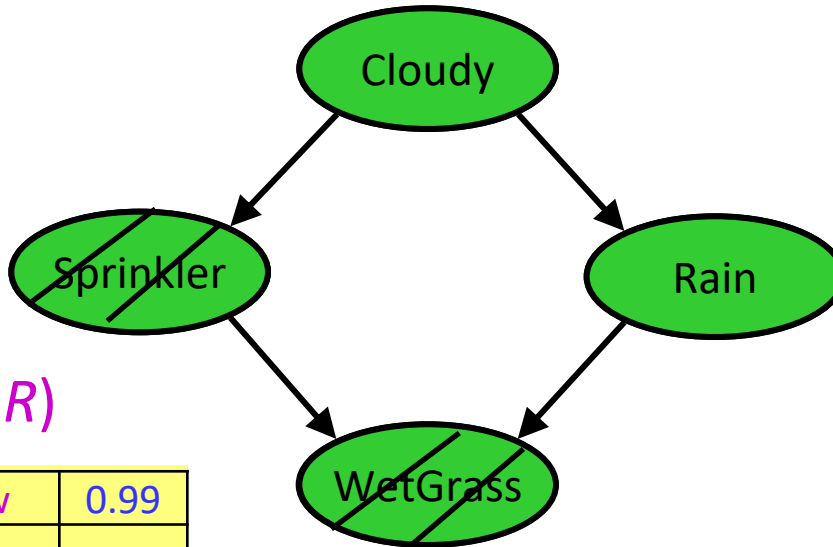
c	s	0.1
	$\neg s$	0.9
$\neg c$	s	0.5
	$\neg s$	0.5

$P(R | C)$

c	r	0.8
	$\neg r$	0.2
$\neg c$	r	0.2
	$\neg r$	0.8

$P(W | S, R)$

s	r	w	0.99
		$\neg w$	0.01
$\neg s$	$\neg r$	w	0.90
		$\neg w$	0.10
	r	w	0.90
		$\neg w$	0.10
$\neg s$	$\neg r$	w	0.01
		$\neg w$	0.99



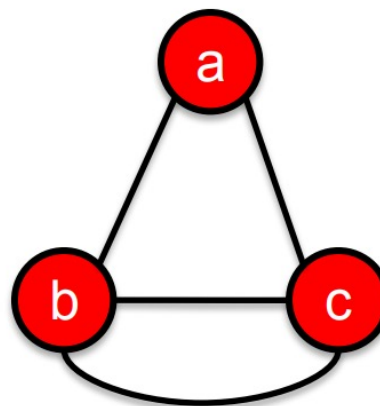
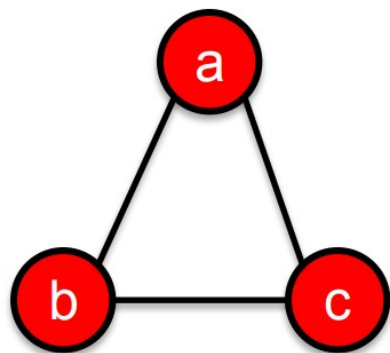
Samples:

c, s, r, w

$w = 1.0 \times 0.1 \times 0.99$

吉布斯采样

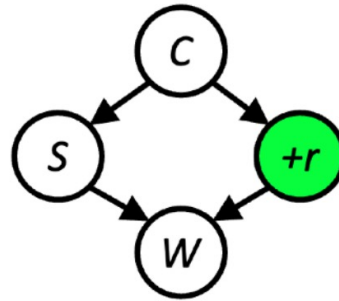
在下列两张图上随机游走 (random walk)，无限次之后，
每个节点停留的次数



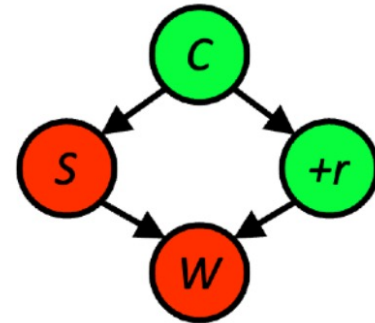
吉布斯采样

- Step 1: 固定观测变量

- $R = +r$

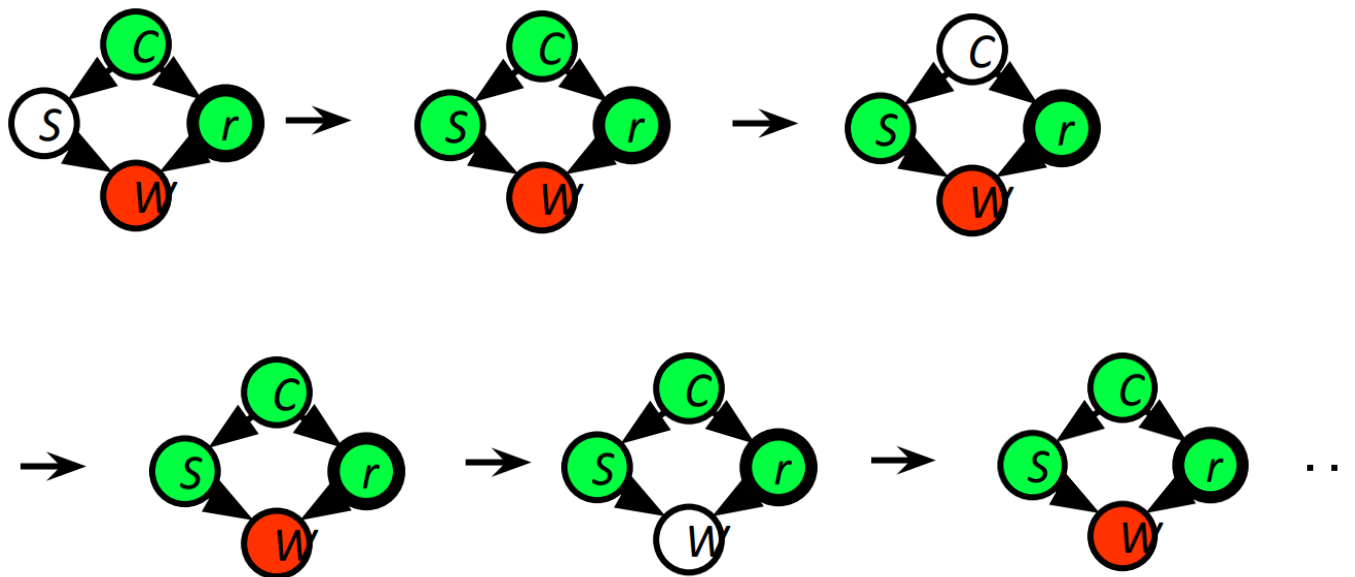


- Step 2: 随机初始化其它变量

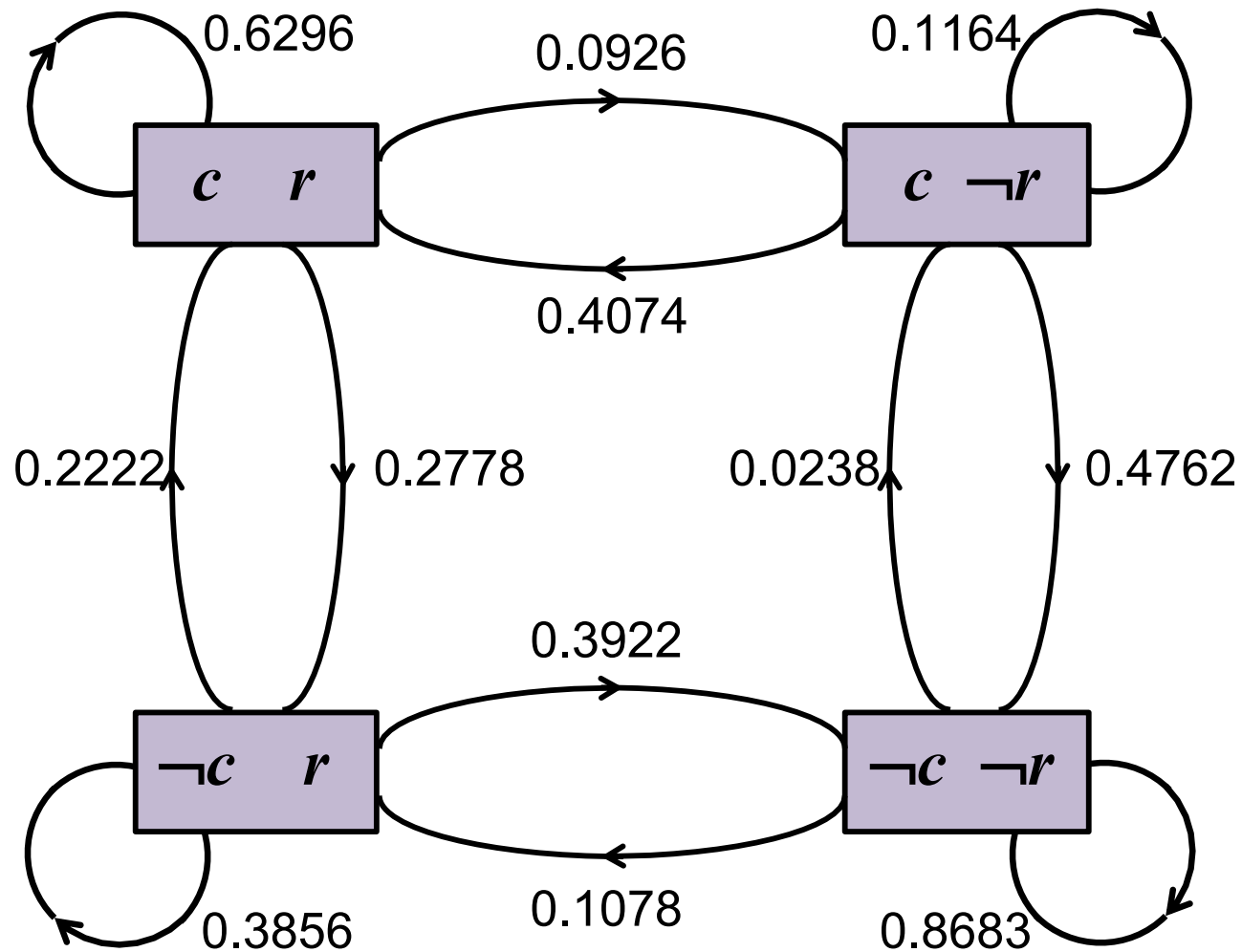


吉布斯采样

- Step 3: 重复
 - 选择一个未观测变量 X
 - Resample X from $P(X|all\ other\ variables)$



Gibbs采样: Given S,W

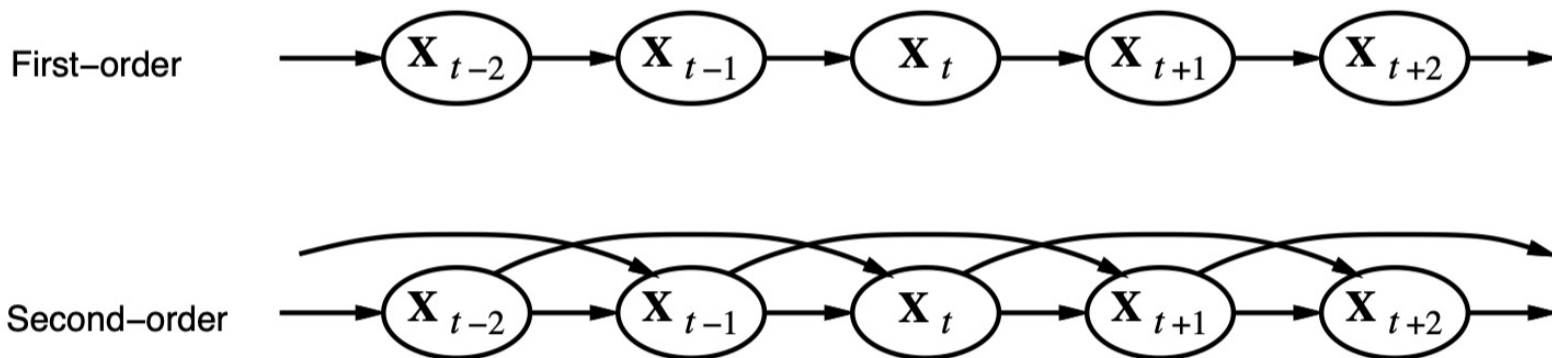


Summary

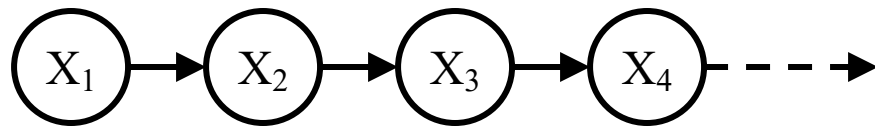
- Prior Sampling P :
 - 基于分布 $P(x_1, \dots, x_n)$ 生成样本
 - Rejection Sampling $P(Q | e)$:
 - 拒绝掉不符合证据 e 的样本
 - Likelihood Weighting $P(Q | e)$:
 - 固定证据变量，对样本进行加权
 - Gibbs sampling $P(Q | e)$:
 - Wander around in e space
 - Average what you see
-

马尔可夫模型(Markov Model)

- **状态变量**：给定时刻 t ，随机变量 X 的取值 $X_1, \dots, X_t$
- **初始状态概率**：随机变量 X 的先验概率分布
- **转移模型(transition model)**：指定世界如何随时间演变，给定过去的状态变量取值之后，确定最新状态的概率分布 $P(X_t | X_1, \dots, X_{t-1})$
- **马尔可夫假设**：当前状态只依赖于有限的固定数量的过去状态



马尔可夫模型(Markov Model)



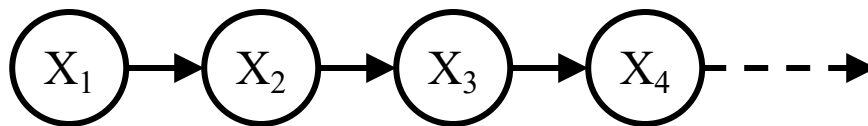
$$P(X_1)$$

$$P(X_t|X_{t-1})$$

- **稳态假设(Stationarity assumption)**: 转移模型不变, 即变化的过程是有本身不随时间变化的规律支配的

注意区分静态和稳态

马尔可夫模型(Markov Model)



$$P(X_1) \quad P(X_t|X_{t-1})$$

- 联合概率分布:

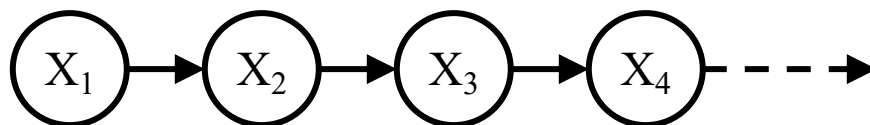
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- 一般化形式:

$$P(X_1, X_2, \dots, X_T) = P(X_1)P(X_2|X_1) \cdots P(X_T|X_{T-1})$$

$$= P(X_1) \prod_{t=2}^T P(X_t|X_{t-1})$$

马尔可夫模型(Markov Model)



- 隐含了什么样的(条件)独立性?

$$X_t \perp X_1, \dots, X_{t-2} | X_{t-1}$$

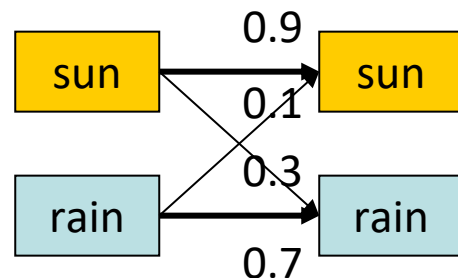
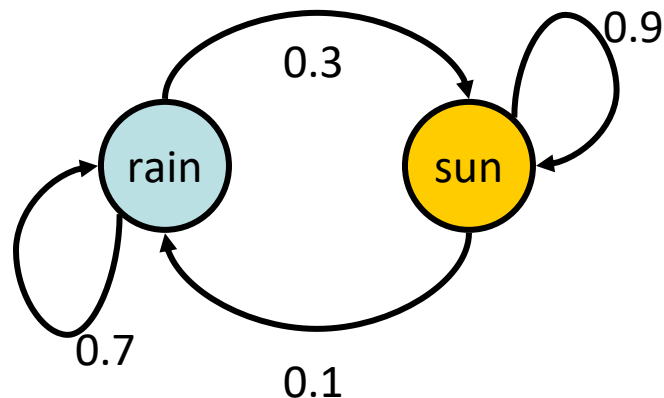
一阶马尔科夫性 $\rightarrow$ 马尔科夫链

马尔可夫模型(Markov Model)

- 状态: $X = \{rain, sun\}$
- 起始分布: $[1.0\ sun, 0.0\ rain]$
- 状态转移: $P(X_t|X_{t-1})$

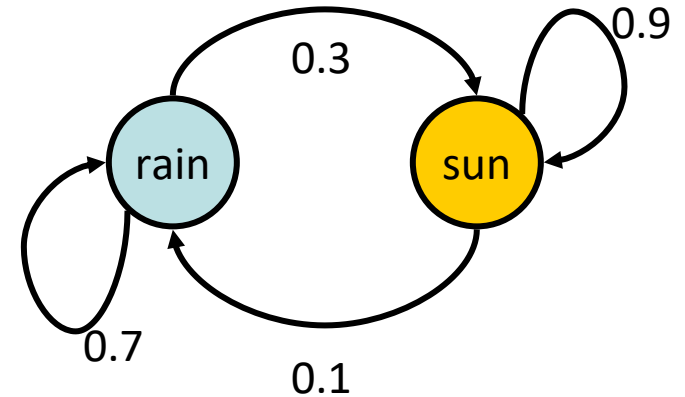
X_{t-1}	X_t	$P(X_t X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

另外两种表示转移模型的方法：
概率转移图和轨迹图



马尔可夫模型(Markov Model)

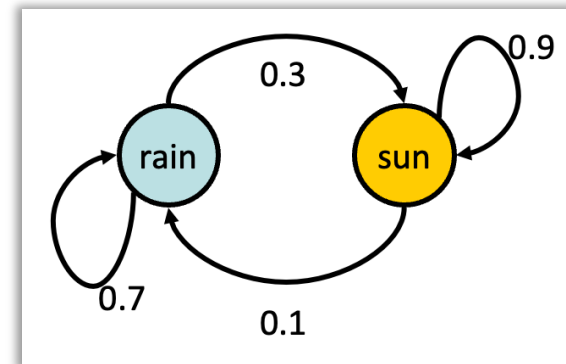
- 起始分布: $[1.0 \text{ sun}, 0.0 \text{ rain}]$



- $P(X_2 = \text{sun})$?

$$\begin{aligned} P(X_2 = \text{sun}) &= P(X_2 = \text{sun} | X_1 = \text{sun})P(X_1 = \text{sun}) + \\ &\quad P(X_2 = \text{sun} | X_1 = \text{rain})P(X_1 = \text{rain}) \\ &= 0.9 * 1.0 + 0.3 * 0.3 = 0.9 \end{aligned}$$

马尔可夫模型(Markov Model)



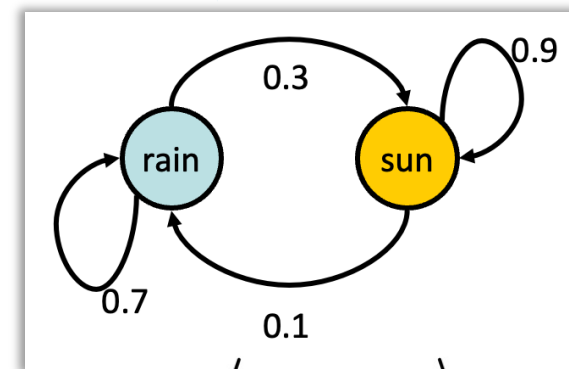
- 起始状态: *sun*

$$\begin{array}{ccccccc} \left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle & \xrightarrow{\hspace{1cm}} & & P(X_\infty) \\ P(X_1) & P(X_2) & P(X_3) & P(X_4) & & & \end{array}$$

- 起始状态: *rain*

$$\begin{array}{ccccccc} \left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle & \xrightarrow{\hspace{1cm}} & & P(X_\infty) \\ P(X_1) & P(X_2) & P(X_3) & P(X_4) & & & \end{array}$$

马尔可夫模型(Markov Model)



- 起始状态: *sun*

$$\begin{array}{ccccccc}
 \left\langle \begin{array}{c} 1.0 \\ 0.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.9 \\ 0.1 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.84 \\ 0.16 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.804 \\ 0.196 \end{array} \right\rangle & \longrightarrow & \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\
 P(X_1) & P(X_2) & P(X_3) & P(X_4) & & P(X_\infty)
 \end{array}$$

- 起始状态: *rain*

$$\begin{array}{ccccccc}
 \left\langle \begin{array}{c} 0.0 \\ 1.0 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.3 \\ 0.7 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.48 \\ 0.52 \end{array} \right\rangle & \left\langle \begin{array}{c} 0.588 \\ 0.412 \end{array} \right\rangle & \longrightarrow & \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle \\
 P(X_1) & P(X_2) & P(X_3) & P(X_4) & & P(X_\infty)
 \end{array}$$

- 任意初始状态

$$\left\langle \begin{array}{c} p \\ 1 - p \end{array} \right\rangle \xrightarrow{\quad \dots \quad} \left\langle \begin{array}{c} 0.75 \\ 0.25 \end{array} \right\rangle$$

$P(X_1)$
 $P(X_\infty)$

马尔可夫模型(Markov Model)

□ 平稳分布：

- 最终到达的分布 P_∞ 称为平稳分布
- 满足如下条件：

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

平稳分布的应用：网页排序PageRank算法

- 假设互联网是一个有向图，在其基础上定义一阶马尔科夫链，表示网页浏览者在互联网上随机浏览网页的过程
 - 转移概率：
 - 以 $1-C$ 的概率，随机跟随页面中的外链跳转，模拟通过点击外链完成页面跳转
 - 以 C 的概率，均匀且随机的跳转到任意其它页面，模拟通过地址栏输入网页地址完成访问，同时保障状态图的全连通性
 - 直观上，如果指向一个网页的超链接越多，随机跳转到该网页的概率也就越高，该网页的PageRank 值就越高，这个网页也就越重要
 - 1996 年由Page 和Brin提出，用于谷歌搜索引擎的网页排序
-

平稳分布的应用：扩散模型

- Text-based image/art generation
 - Dall-E 2, Imagen, Stable Diffusion, MidJourney...



[Christian Beltrami/MidJourney]



vibrant portrait painting of Salvador Dali with a robotic half face



a shiba inu wearing a beret and black turtleneck



a close up of a handpalm with leaves growing from it



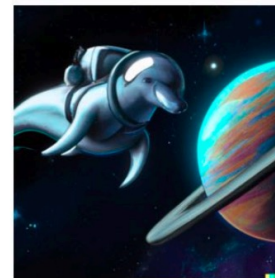
an espresso machine that makes coffee from human souls, artstation



panda mad scientist mixing sparkling chemicals, artstation



a corgi's head depicted as an explosion of a nebula



a dolphin in an astronaut suit on saturn, artstation



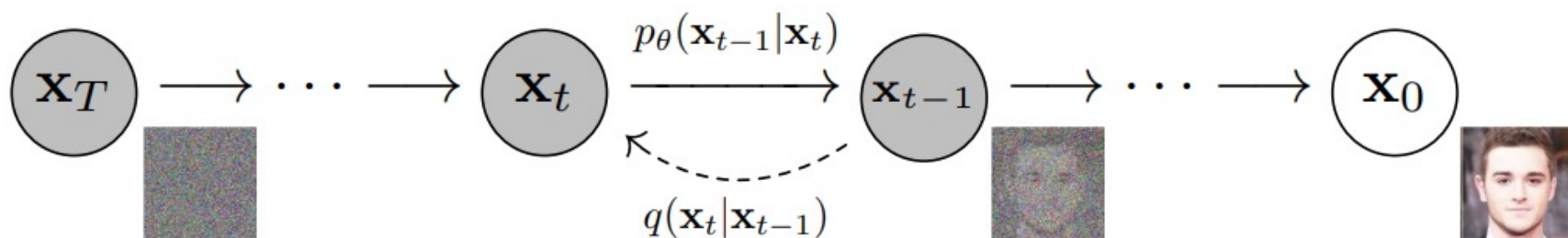
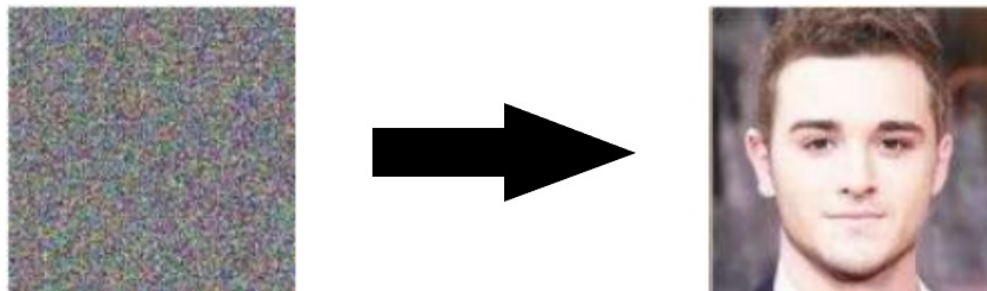
a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese



a teddybear on a skateboard in times square

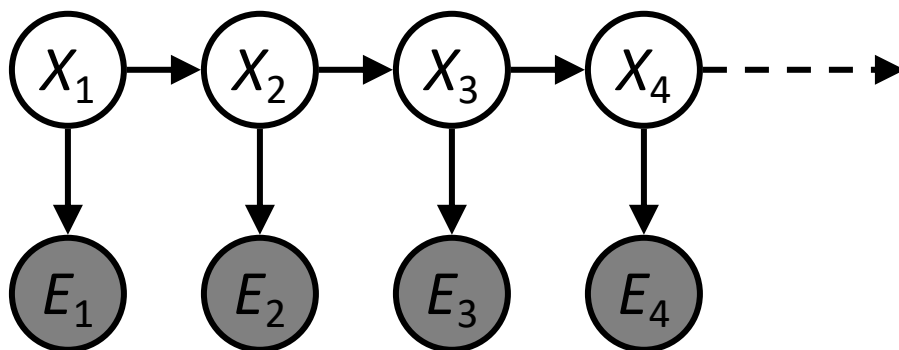
Figure 1: Selected 1024×1024 samples from a production version of our model. [OpenAI]

平稳分布的应用：扩散模型

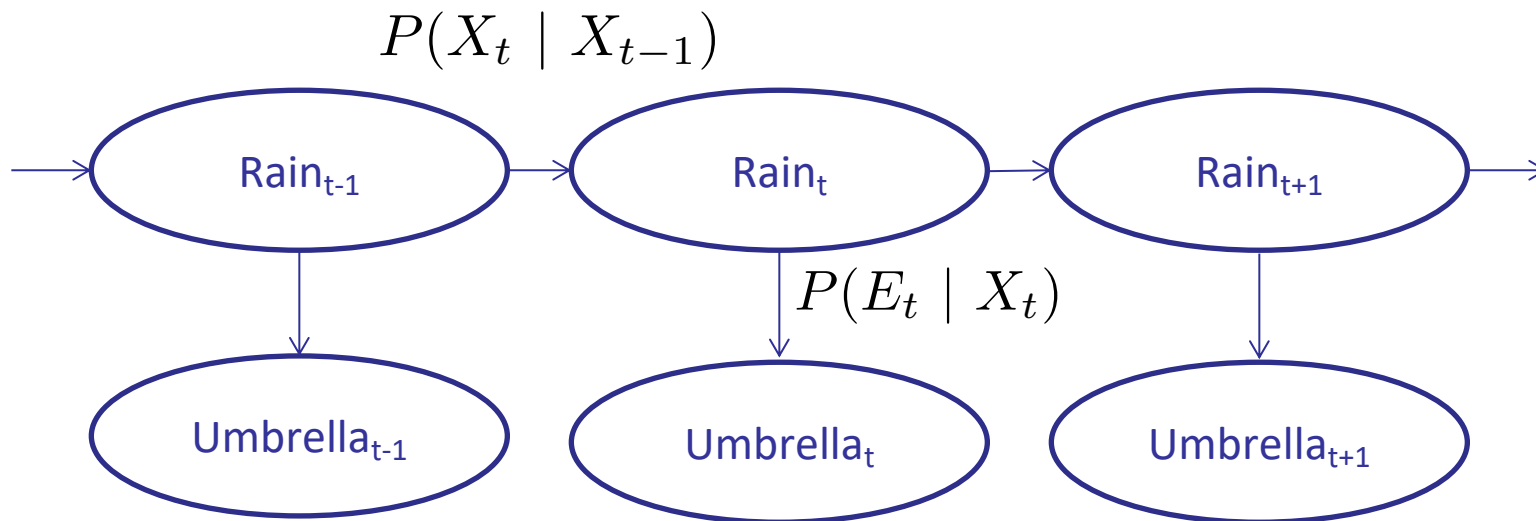


隐马尔可夫模型(Hidden Markov Models)

- 隐马尔可夫模型(HMMs):
 - 由一个隐藏的马马尔可夫链随机生成不可观测的状态序列(state sequence), 再由各个状态生成一个观测从而产生观测序列(observation sequence)的过程



隐马尔可夫模型(Hidden Markov Models)



- 初始概率: $P(X_1)$
- 转移概率: $P(X_t | X_{t-1})$
- 观测(发射)概率: $P(E_t | X_t)$

R_{t-1}	$P(R_t R_{t-1})$
+r	0.7
-r	0.3

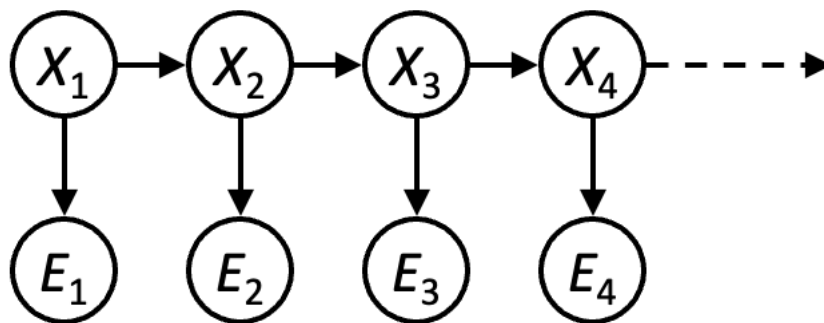
R_t	$P(U_t R_t)$
+r	0.9
-r	0.2

HMM的三要素

隐马尔可夫模型(Hidden Markov Models)

- HMM的两条独立性假设：

- 马尔可夫假设：隐藏的马尔可夫链在任意时刻 t 的状态只依赖其前一时刻的状态
- 观测独立性假设：任意时刻的观测只依赖于该时刻的马尔可夫链状态



$$P(X_1, E_1, \dots, X_T, E_T) = P(X_1)P(E_1|X_1) \prod_{t=2}^T P(X_t|X_{t-1})P(E_t|X_t)$$

隐马尔可夫模型(Hidden Markov Models)

$$P(X_1, E_1, \dots, X_T, E_T) = P(X_1)P(E_1|X_1) \prod_{t=2}^T P(X_t|X_{t-1})P(E_t|X_t)$$

- 对于所有时刻 t ，给定前一时刻隐状态，当前隐状态独立于所有过去的隐状态和观测状态

$$X_t \perp X_1, E_1 \dots, X_{t-2}, E_{t-2}, E_{t-1} | X_{t-1}$$

- 给定当前时刻隐状态，当前时刻的可见状态独立于过去所有的隐状态和可见状态

$$E_t \perp X_1, E_1 \dots, X_{t-2}, E_{t-2}, E_{t-1} | X_t$$

隐马尔可夫模型的应用

- 语音识别
 - 观测序列为声学信号
 - 状态序列为语言文字
 - 机器翻译
 - 观测序列为源语言文字
 - 状态序列为目标语言文字
 - 机器人跟踪定位
 - 观测序列为连续的传感器信号
 - 状态序列为机器人所处的位置
-