概率图模型理论及应用

Theory and Applications of Probabilistic Graphical Models

(Lesson 2)

欧智坚

清华大学电子工程系

Addr: 罗姆楼 6-104

Tel: 62796193

Email: ozj@tsinghua.edu.cn

考核方式

- 课后作业 (4次): 20%
  - 掌握基本概念和原理
  - 按时交至网络学堂，迟交分数打九折。
- 阅读及笔记（scribe）一次: 20%
  - 13节课，课前微信群提交阅读摘要（1页）：总结，提出问题；
  - 课堂笔记；
  - 课后扩展阅读和调研，综合你的阅读和课堂笔记，形成一节资料，当次课一周后（周日23:59）email交初稿，两周后（周日23:59） email交终稿，Latex模板。
  - 若成组(<=3人)，按组给分，说明分工
- 大作业 (自行选题的project): 60%
  - Applying 应用解决实际问题
  - Theoretic 偏理论探索
  - 第17周报告
  - 若成组(<=2人)，按组给分，说明分工

课程章节

- 第一章 引言（1）
- 第二章 图模型的表示理论（2）
  - **Semantics (DGM, UGM)**
  - HMM, CRF
- 第三章 图模型的推理理论（6）
  - 精确推理：variable-elimination，cluster-tree，triangulate
  - 连续变量：Kalman
  - 采样近似：sampling
  - 变分近似：variational
- 第四章 图模型的学习理论（3）
  - 参数学习：maxlikelihoodEstimate，UGMLearning, BayesEstimate
  - 结构学习：StructureLearning
- 第五章 一个综合例子（1）

同学课前阅读

问题

1、贝叶斯网和马尔可夫网各有什么优势和劣势，分别适合建模哪些类别的问题？

2、计算机视觉中常用 CRF 对问题建模，很多情况下，CRF 模型其实就是数据项 + 平滑项，最优化的方法往往也采用非线性最小二乘的方法（Gauss-Newton 或 L-M），似乎没有学过图模型，也可以解决 CRF 问题。请问概率图模型到底给了计算机视觉中的 CRF 问题哪些新的东西？是否是更加高效的优化方法？

Why Graphical Model？

- 丰富的模型表达能力（Representation）
  - **统一**了目前广泛应用的许多统计模型和方法
  - 通过 画图 以建模；通过 读图 以分析变量间关系
- 强大的推理计算能力（Inference）
  - 原理性算法，通用性，一般性
  - 面对具体的新模型，运用原理而不必自己重新设计推理算法

求：$p(S=1|W=1)=?$ $p(R=1|W=1)=?$

- 全面的学习理论（Learning）
  - 结构学习
  - 拍脑袋＋让数据来说话：领域知识与样本信息（数据）有机结合

View of a family of distributions 分布族的观点

- 一个图g的分布，一般而言，不是指一个具体的分布，而是指一族分布
  - 穷尽 x_v 的形式、每个结点处的局部条件分布$\{\ p(x_v \mid x_{pa(v)})\ \}$的可能形式，得到一个分布的集合，记为 $\mathbb{M}(g)$

$$\mathbb{M}(g) = \left\{ p(x_V) : p(x_V) = \prod_{v \in V} p\left(x_v \mid x_{pa(v)}\right) \right\}$$

$$p(V) = \prod_{v \in V} p\left(v \mid pa(v)\right)$$

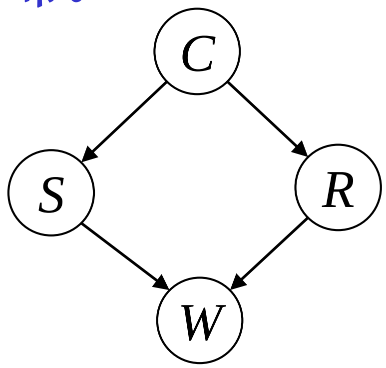
  - **一般讨论：**对这一族分布中的每个分布都成立的结论
  - **从一般到具体，可以有不同抽象层次看待一个图的分布。**

$$p(C,R,S,W) = p(C)\,p(R \mid C)\,p(S \mid C)\,p(W \mid R,S)$$

$$R \perp S \mid C$$

$$W \perp C \mid S, R$$

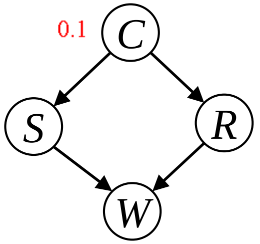
C	$p(S=0 \mid C)$	$p(S=1 \mid C)$
0	0.5	0.5
1	0.9	0.1



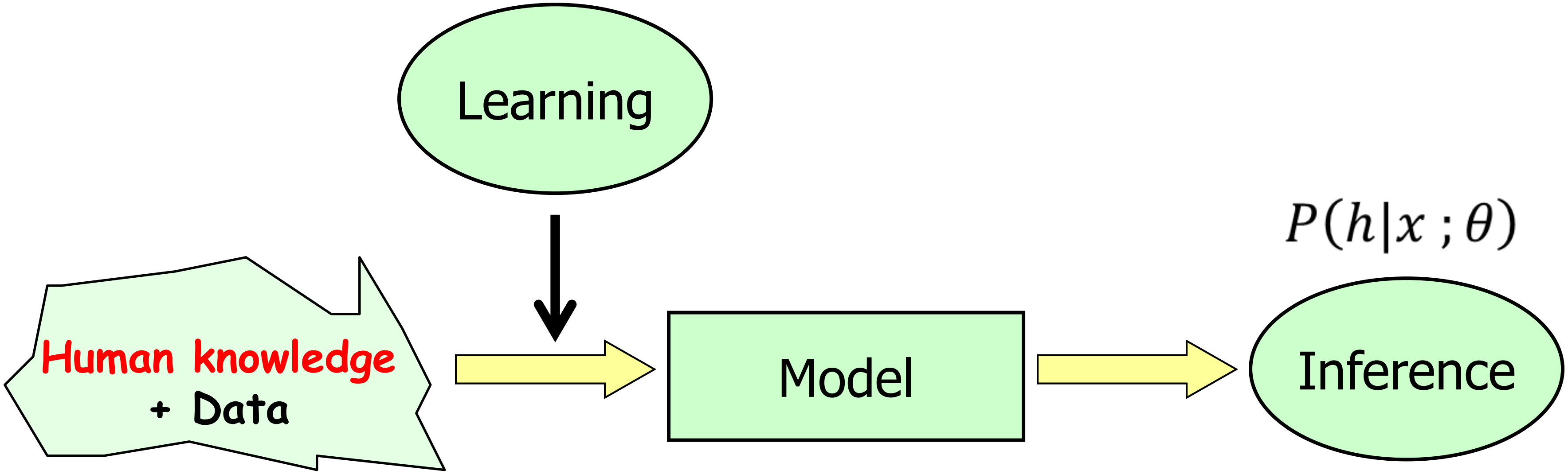
图模型建模的四要素

- 当使用图模型去表示一个具体的概率分布时（ to describe a particular distribution），需逐步明确以下四要素
- 语义（Semantics）
  - 定义了图和概率分布如何发生联系（有向图、无向图、因子图、链图、...）
- 结构（Structure）
  - 定义随机变量及之间联系（即明确图包含哪些结点以及边）
- 实现（Implementation）
  - 指定结点类型（ discrete/continuous ）以及局部函数的具体形式
- 参数（Parameter）
  - 凭经验指定或利用数据进行估计——局部函数的待定参数的取值

C	$p(S=0\|C)$	$p(S=1\|C)$
0	0.5	0.5
1	0.9	0.1



图模型理论的三要素



$P(x, h\,;\ \theta)$: Generative model,

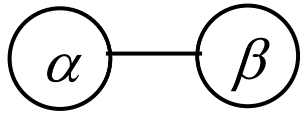
$e.g. GMMs, HMMs, FA, MRFs, LDA,\ \ldots$

$P(h|x\,;\theta)$: Discriminative model,

$e.g. Logistic\ Regression, CRFs, NNs,\ \ldots$

Graph Theory

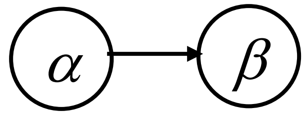
Basic concepts – undirected edge, neighbor



we have an undirected edge between α and β	$\alpha \sim \beta$	If both ordered-pairs (α, β) and (β, α) belong to E
we also say that α and β are neighbors, α is a neighbour of β, or β is a neighbor of α		

the set of neighbors of β 邻居集	$ne(\beta)$	$ne(\beta) = \{\alpha : \alpha \sim \beta\}$

Basic concepts – directed edge, parent, child



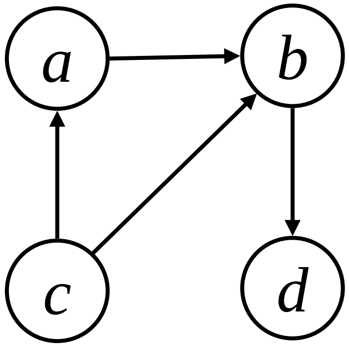
we have an directed edge from α to β	$\alpha \rightarrow \beta$	If $(\alpha, \beta) \in E$ but $(\beta, \alpha) \notin E$
we also say that α is a parent of β , and that β is a child of α		

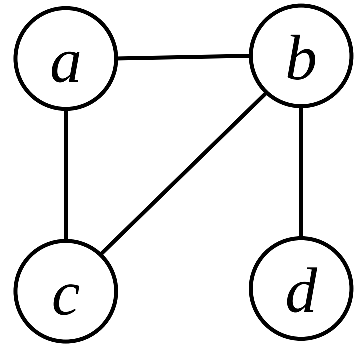
the set of parents of β 父结点集	$pa(\beta)$	$pa(\beta) = \{\alpha : \alpha \rightarrow \beta\}$
the set of children of α 子结点集	$ch(\alpha)$	$ch(\alpha) = \{\beta : \alpha \rightarrow \beta\}$

Basic concepts – un/directed graph

we say D is a directed graph		If *all* the edges of a graph D are directed
we say g is a undirected graph		If *all* the edges of a graph g are undirected

Basic concepts – family; boundary, closure





有向图

the family of β 家庭	$fa(\beta)$	$fa(\beta)=\{\beta\}\cup pa(\beta)$

无向图

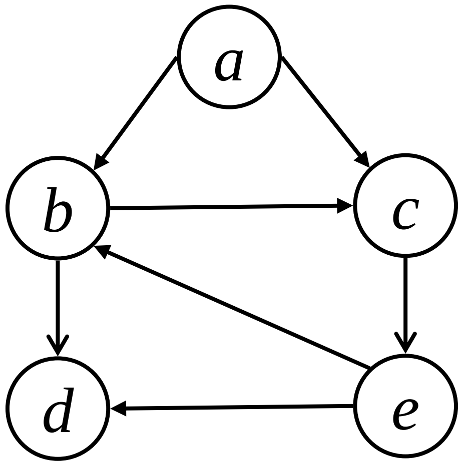
Boundary of node α 边界	$bd(\alpha)$	$bd(\alpha)=ne(\alpha)$
Closure of node α 闭包	$cl(\alpha)$	$cl(\alpha)=bd(\alpha)\cup\{\alpha\}$

Basic concepts – path, trail

path 路径		从α_0到α_n的一条长为n的路径，是指存在n+1个不同结点构成的序列$\alpha_0,\ldots,\ \alpha_n$， 且 $(\alpha_{i-1},\ \alpha_i)\in E$，$i = 1,\ \ldots,\ n$.

- 一条路径不会和自己相交；不能逆着有向边的方向而行

we say that α leads to β α 可到达 β	$\alpha \mapsto \beta$	如果存在一条从α 到 β 的路径



Basic concepts – cycle, DAG, trail

n-cycle 环		从α_1到α_n的一条长为n的环，是指存在n个不同结点构成的序列$\alpha_1,\ldots,\alpha_n$，且$(\alpha_i,\alpha_{i+1})\in E$，$i=1,\ldots,n\text{-}1$，$(\alpha_n,\alpha_1)\in E$

acyclic graph		If it does not possess any cycles
directed acyclic graph (DAG)		A directed graph which is acyclic

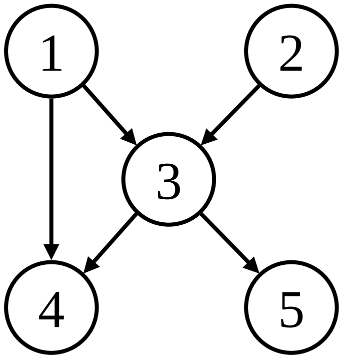
trail from α to β 迹		从α_0到α_n的一条长为n的迹，是指存在n+1个不同结点构成的序列$\alpha_0,\ldots,\alpha_n$，且 $\alpha_{i\text{-}1}\rightarrow\alpha_i$, or $\alpha_i\rightarrow\alpha_{i\text{-}1}$, or $\alpha_{i\text{-}1}\sim\alpha_i$，$i=1,\ldots,n$.

- 一条迹不会和自己相交；可以逆着有向边的方向而行

Basic concepts – well-order of DAG

well-order of a DAG (Directed Acyclic Graph) 有向无环图的良序		我们总可以对一个DAG的结点进行排序/编号，使得：图中的有向边总是从低编号的结点指向高编号的结点。这样一种排序称为良序

predecessors of node α	$pr(\alpha)$	在DAG的一个良序中，结点编号比 α 小的结点集合称为 α 的前辈结点集

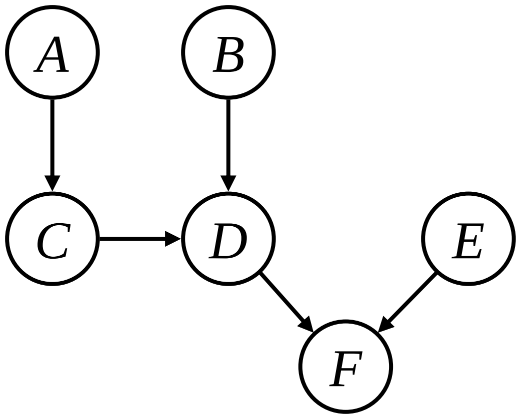


Here is a well-ordering of the nodes in the DAG

Basic concepts – ancestors, non/descendants

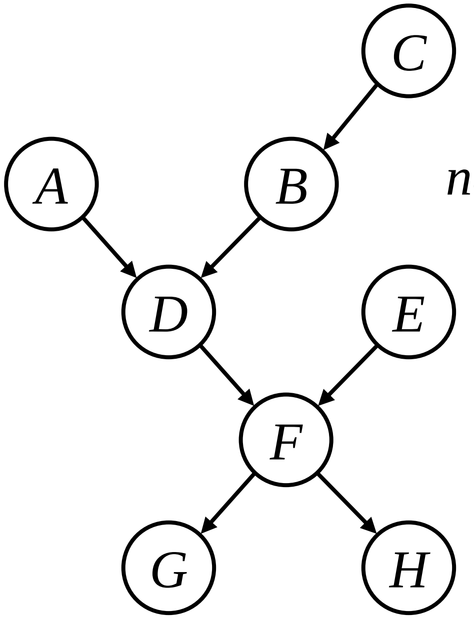
有向图

ancestors (祖先) of node β	$an(\beta)$	$an(\beta)=\{\alpha:\alpha\mapsto\beta \text{ but not } \beta\mapsto\alpha\}$
descendants (后代) of α	$de(\alpha)$	$de(\alpha)=\{\beta:\alpha\mapsto\beta \text{ but not } \beta\mapsto\alpha\}$
non-descendants (非后代) of α	$nd(\alpha)$	$nd(\alpha)=V\setminus(\alpha\cup de(\alpha))$



$an(C)$ =

$de(C)$ =



$nd(D)$ =

Conditional independence (CI)

Definition

- $X \perp Y \mid Z$
  if and only if the following equivalent condition holds
  - $p(x \mid y, z) \equiv p(x \mid z)$ if $p(y, z) > 0$
  - $p(x \mid y, z)$ has the form $a(x, z)$ if $p(y, z) > 0$
  - $p(x, y \mid z) \equiv p(x \mid z)\, p(y \mid z)$ if $p(z) > 0$
  - $p(x, y \mid z)$ has the form $a(x, z)\, b(y, z)$ if $p(z) > 0$
  - $p(x, y, z) \equiv p(x \mid z)\, p(y \mid z)\, p(z)$
  - $p(x, y, z) \equiv p(x, z)\, p(y, z) / p(z)$ if $p(z) > 0$
  - $p(x, y, z)$ has the form $a(x, z)\, b(y, z)$

Properties of CI—条件独立性的运算法则

- The ternary relation $X \perp Y \mid Z$ has the following properties:
  - P1: Symmetry
    If $X \perp Y \mid Z$ then $Y \perp X \mid Z$
  - P2: Decomposition
    If $X \perp Y \mid Z$ and U is a function of X then $U \perp Y \mid Z$
  - P3: Weak Union
    If $X \perp Y \mid Z$ and U is a function of X then $X \perp Y \mid (Z, U)$
  - P4: Contraction
    If $X \perp Y \mid W$ and $X \perp Z \mid (W, Y)$ then $X \perp (Y, Z) \mid W$
  - P5: Intersection (hold under certain conditions)
    If $X \perp Y \mid (Z, W)$ and $X \perp Z \mid (Y, W)$ then $X \perp (Y, Z) \mid W$

Properties of CI - Example of P5 fails

- $X \perp Y \mid Z$ and $X \perp Z \mid Y \Rightarrow ? \; X \perp (Y, Z)$
  - Define: $p(X{=}Y{=}Z{=}1) = p(X{=}Y{=}Z{=}0) = 0.5$

    Other configurations of X, Y and Z have 0 probability.
  - $$p(x \mid z) = \frac{p(x,z)}{p(z)} = \frac{0.5 \times \delta(x=z)}{0.5} = \delta(x=z) \qquad p(y \mid z) = \delta(y=z)$$

    $$p(x,y \mid z) = \frac{p(x,y,z)}{p(z)} = \frac{0.5 \times \delta(x=y=z)}{0.5} = \delta(x=y=z)$$

    $$= \delta(x=z)\,\delta(y=z) = p(x \mid z)\,p(y \mid z)$$
  - This gives $X \perp Y \mid Z$. Similarly, $X \perp Z \mid Y$.
  - $$p(x \mid y=1, z=1) = \frac{p(x, y=1, z=1)}{p(y=1, z=1)} = \frac{0.5 \times \delta(x=1)}{0.5} = \delta(x=1)$$

    $$p(x=0) = p(x=1) = 0.5$$
  - $X \perp (Y, Z)$ is not true. Given (Y, Z), we know $X{=}Y{=}Z$.

Properties of CI - P5

- P5 is true when the distribution p is positive.

  If $\begin{array}{l} X \perp Y \mid (Z,W) \\ X \perp Z \mid (Y,W) \end{array}$ and $p(x,y,z,w) > 0, \forall x,y,z,w$

  then $X \perp (Y,Z) \mid W$

- Proof: $p(x,y,z,w) = g(x,z,w)h(y,z,w)$

$$= k(x,y,w)l(y,z,w)$$

for suitable strictly positive functions g, h, k, l.

Thus, for all z we must have $k(x,y,w) = \dfrac{g(x,z,w)h(y,z,w)}{l(y,z,w)}$

Hence, choosing a fixed $z = z_0$, we have

$$k(x,y,w) = g(x,z_0,w)\frac{h(y,z_0,w)}{l(y,z_0,w)} = \pi(x,w)\rho(y,w)$$

Thus, $p(x,y,z,w) = \pi(x,w)\rho(y,w) \times l(y,z,w) \Rightarrow X \perp (Y,Z) \mid W$

第二章 图模型的表示理论

语 义

一个图表示了怎样的概率分布

图 $\rightleftarrows$ 概率分布

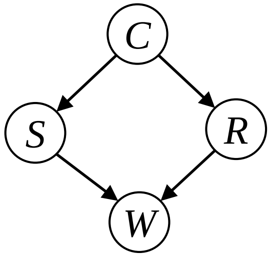
一个概率分布如何表示成一个图

Definition - (DF) property

赋予概率分布一新表述：先定义性质，性质诱导分布

❖ 称一个分布 $p(x_V)$ 服从 依图D的有向分解性，如果

$$p\left(x_V\right)=\prod_{v\in V}p\left(x_v\mid x_{pa(v)}\right)$$



由依图D的有向分解性诱导了一族分布：$\mathbb{M}_{DF}(D)$

$$\mathbb{M}_{DF}\left(D\right)=\left\{p(x_V):p(x_V)=\prod_{v\in V}p\left(x_v\mid x_{pa(v)}\right)\right\}$$

Definition - (DO) property

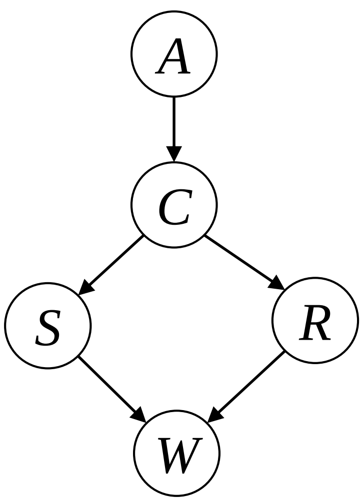
称一个分布 $p(x_V)$ 服从依图 D 的有向有序Markov性，

in some given well-ordering of the nodes (CDLS p73)

如果在图D的**某个**良序下，成立

$$v \perp pr(v) \setminus pa(v) \mid pa(v)$$

由依图D的有向有序Markov性诱导了一族分布: $\mathbb{M}_{DO}(D)$



$p(A, C, S, R, W)$

在**某个**良序下，成立 $v \perp pr(v) \setminus pa(v) \mid pa(v)$

$\Downarrow$ 对图的结点数运用数学归纳法

$$p(V) = \prod_{v \in V} p(v \mid pa(v))$$

$\Downarrow$

在**任意**良序下，成立 $v \perp pr(v) \setminus pa(v) \mid pa(v)$

不同良序下写出的CI是等价的，统称为DO

View of a family of distributions

依图D的 有向分解性	$\rightleftarrows$	依图D的 有向有序Markov性
$p(V)=\prod_{v\in V} p(v \mid pa(v))$		$v \perp pr(v) \setminus pa(v) \mid pa(v)$

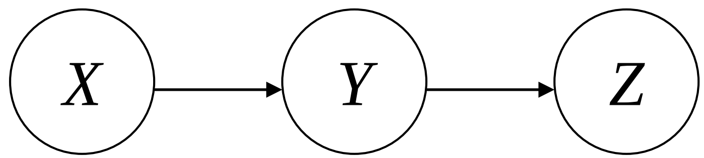
- Theorem: $\mathbb{M}_{DF}(D)$ = $\mathbb{M}_{DO}(D)$，合记为$\mathbb{M}(D)$

At the core of the graphical model representation

- 在建模和算法分析中，灵活地解读一个概率图

Three canonical DAG's - Serial

- Serial connection (Markov Chain)



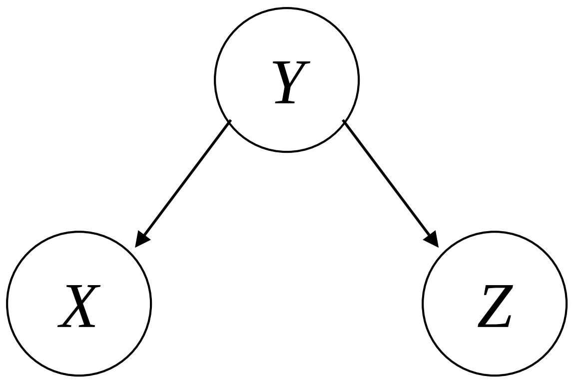
$$p(x,y,z)=p(x)\,p(y\mid x)\,p(z\mid y)$$

$$X \perp Z \mid Y$$

- 当几个随机现象串行影响时，给定当前现象 Z 的直接原因 Y，则当前现象 Z 的发生与更早的原因 X 无关。

Three canonical DAG's - Diverging

- Diverging connection



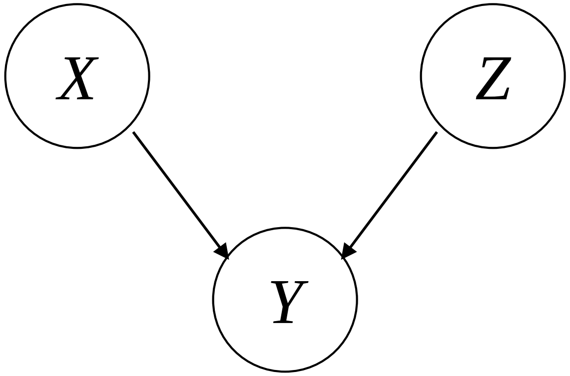
$$p(x,y,z) = p(x|y)\,p(y)\,p(z|y)$$

$$X \perp Z \mid Y$$

- 当一个随机现象 Y 是另两个随机现象 X, Z 的共同原因时，给定原因Y，则这两个随机现象 X, Z 彼此独立。

Three canonical DAG's - Converging

- Converging connection



$$p(x, y, z) = p(x)\, p(y \mid x, z)\, p(z)$$

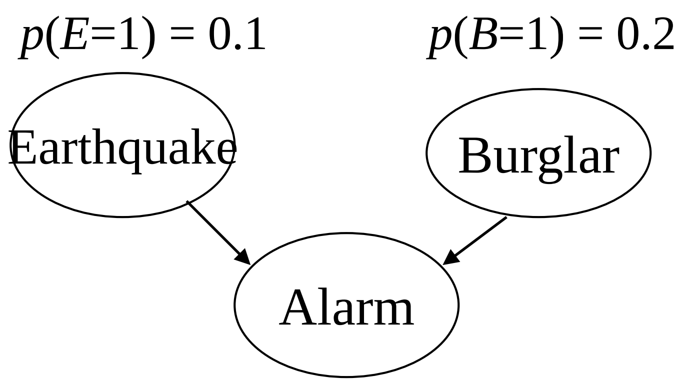
$$X \perp Z$$

$$X \perp Z \mid Y \quad \times$$

Three canonical DAG's - Converging

- "Multiple, competing explanation" scenario

E	B	$p(A=1\|E, B)$	$p(A=0\|E, B)$
0	0	0.1	0.9
1	0	0.8	0.2
0	1	0.8	0.2
1	1	0.99	0.01



- $p(B=1 \mid A=1)$ **?** $p(B=1 \mid A=1, E=1)$ $\qquad E \perp B \mid A$ ×

- 当一个随机现象有两个竞争的原因，其中一个原因发生条件下，另一个原因发生的可能性减少（ explain away ）。

DAG的语义

一个图表示了怎样的概率分布

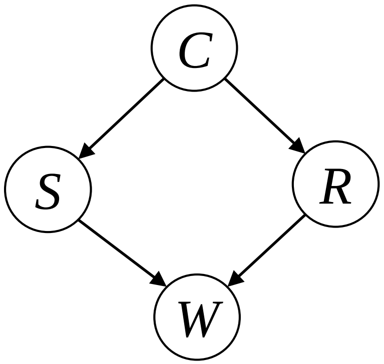
图 ⇄ 概率分布

一个概率分布如何表示成一个图

一个DAG表示一个怎样的概率分布（by definition）

- 满足有向分解性
- 满足有向有序Markov性

一个概率分布如何表示成一个DAG



$$p(C,R,S,W)=p(C)\,p(R\,|\,C)\,p(S\,|\,C)\,p(W\,|\,R,S)$$

I-map (independency map)

- 将一个分布 $p(.)$ 所蕴含/满足/表示的 CI 记为 $\mathbb{I}(p)$

$$p(C,R,S,W)=p(C)p(R\mid C)p(S\mid C)p(W\mid R,S)$$

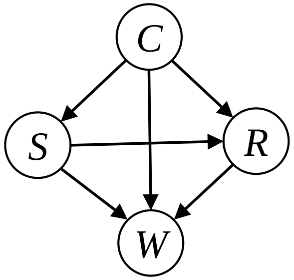
$$\left\{\begin{matrix} R\perp S\mid C \\ W\perp C\mid S,R \\ \dots \end{matrix}\right\}$$

称一个BN D 是一个分布 p 的 I-map，如果

$$\mathbb{I}(\mathbb{M}(D))\subseteq\mathbb{I}(p)$$

这时，我们也称分布 p 可以表示成 图D 。

- 相对分布 p 所表示的条件独立性，图 D 表示了较少的条件独立性。
- 分布 p 可以有一些条件独立性没有反映在图 D 中。
- 一个完备的有向无环图 是 任何分布的 I-map。



称一个BN D 是一个分布 p 的 完美贝叶斯网络表示 (P-map)，如果

$$\mathbb{I}(\mathbb{M}(D))=\mathbb{I}(p)$$

P-map的存在性

- 不是每一个分布都有 完美贝叶斯网络表示。

例子（KF p.81 example 3.6）

- 考虑三个布尔随机变量X, Y, Z的如下联合分布

$$p(x,y,z)=\begin{cases}1/12 & x\oplus y\oplus z=0\\ 1/6 & x\oplus y\oplus z=1\end{cases}$$

x	y	$x\oplus y$
0	0	0
0	1	1
1	0	1
1	1	0

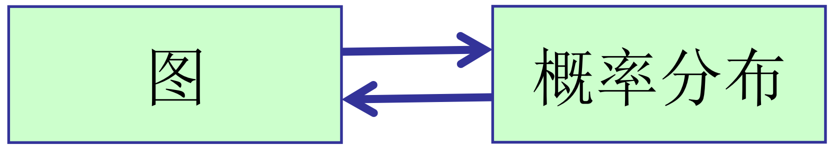
$$(X\perp Y)\in\mathbb{I}(p)$$
$$(X\perp Z)\in\mathbb{I}(p)$$
$$(Y\perp Z)\in\mathbb{I}(p)$$
$$(X\perp Y\mid Z)\notin\mathbb{I}(p)$$
$$(XY\perp Z)\notin\mathbb{I}(p)$$

$$\mathbb{I}\left(\begin{tikzcd} X \arrow[dr] & & Y \arrow[dl] \\ & Z & \end{tikzcd}\right)\overset{?}{=}\mathbb{I}(p)$$

$$\mathbb{I}\left(\begin{tikzcd} X & & Y \\ & Z & \end{tikzcd}\right)\overset{?}{=}\mathbb{I}(p)$$

DAG的语义

一个图表示了怎样的概率分布



一个概率分布如何表示成一个图

一个DAG表示一个怎样的概率分布（by definition）

- 满足有向分解性
- 满足有向有序Markov性

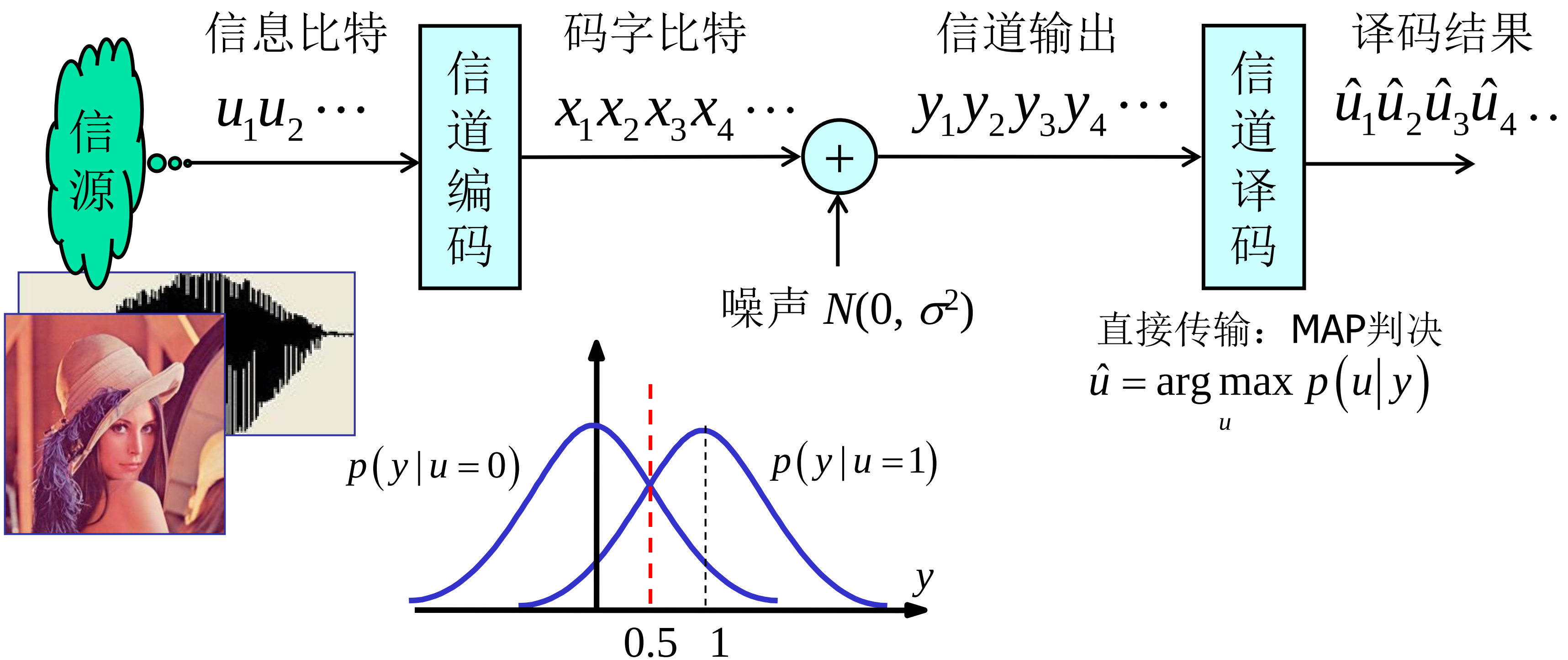
一个概率分布如何表示成一个DAG

- 从条件独立性的包含关系上，寻找尽可能紧凑的I-map

- DAG适用于随机因素间具有明确依赖、影响关系，通过条件分布来表达的建模问题
  - 通过画图来建模

信道译码应用举例

信道编译码



- 引入信道编码：将信息比特 $u_{1:K}$ 为一组编码成码字比特 $x_{1:N}$
- 简单重复编码：K=1，一个信息比特重复发两份，利用 $\dfrac{y_{2k}+y_{2k+1}}{2}$ 判决
- 传输可靠性 与 传输效率

4/7 Hamming code

- 为信息比特加上奇偶校验位（parity-check bits）形成码字
  - K=4, N=7 Hamming code

$$(x_1,x_2,x_3,x_4,x_5,x_6,x_7) \quad \text{生成矩阵 } G$$

$$=(u_1,u_2,u_3,u_4)\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

$$\Rightarrow$$

校验矩阵 H

$$(x_1,x_2,x_3,x_4,x_5,x_6,x_7)\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}=(0,0,0)$$

Hard decoding

- 对信道输出的前4位 $y_{1:4}$ 进行门限为0.5的硬判决，得到 $\hat{u}_{1:4}$

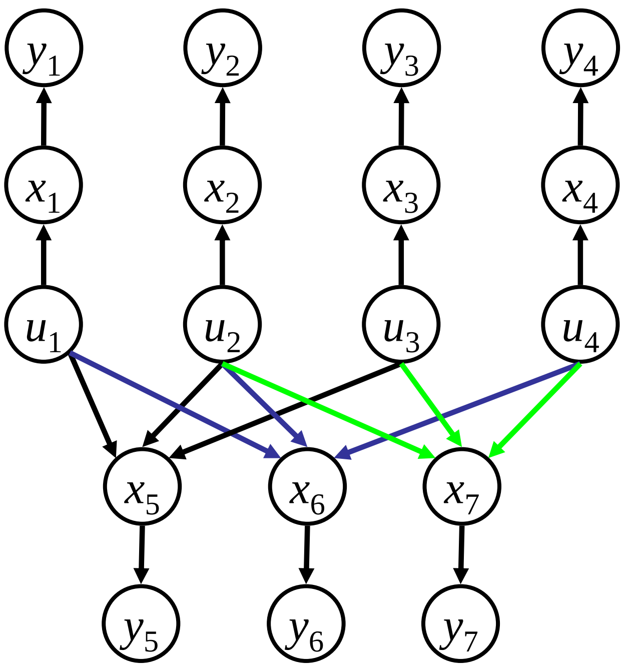
Algebraic decoding

- 对信道输出 $y_{1:7}$ 做硬判决，进行伴随式译码 $\hat{y}_{1:7}\cdot H=0$ ？

Bayes net for a 4/7 Hamming code

(Exact) Probabilistic decoding： 基于 $u_{1:4}$ $x_{1:7}$ $y_{1:7}$的联合分布

$$\hat{u}_k = \arg\max_{u_k=0,1} p\left(u_k \mid y_{1:7}\right)$$



- $p(u_k=0)= p(u_k=1)=0.5$ for $k=1,2,3,4$.
- $p(x_1 \mid u_1) = \delta(x_1 , u_1)$

  $p(x_2 \mid u_2) = \delta(x_2 , u_2)$

  $p(x_3 \mid u_3) = \delta(x_3 , u_3)$

  $p(x_4 \mid u_4) = \delta(x_4 , u_4)$

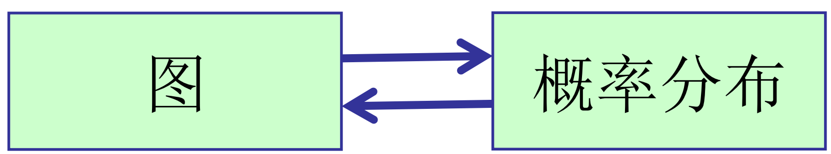
$$p\left(x_5 \mid u_1,u_2,u_3\right)=\delta\left(x_5,u_1+u_2+u_3\right)$$
$$p\left(x_6 \mid u_1,u_2,u_4\right)=\delta\left(x_6,u_1+u_2+u_4\right)$$
$$p\left(x_7 \mid u_2,u_3,u_4\right)=\delta\left(x_7,u_2+u_3+u_4\right)$$

- $p\left(y_n \mid x_n\right)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{\left(y_n-x_n\right)^2}{2\sigma^2}\right\}$ for $n=1,\ldots,7$

第二章 图模型的表示理论

无向图模型

一个图表示了怎样的概率分布



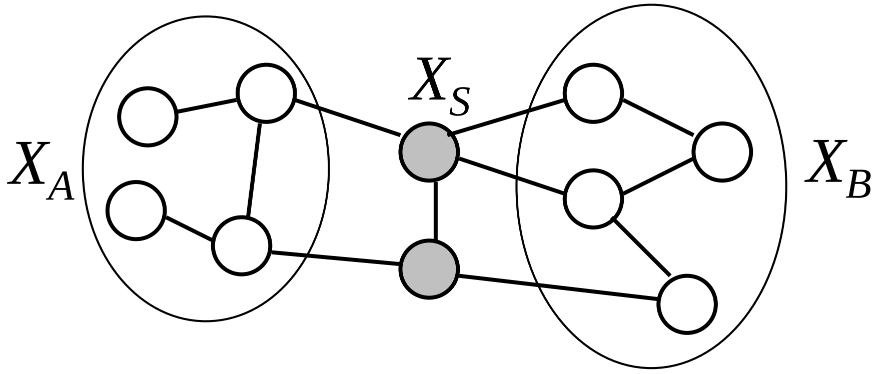
一个概率分布如何表示成一个图

Definition - (G) property

称一个分布 $p(x_V)$ 服从依图g的全局Markov性，如果对任意互不相交的三个结点子集 (A, B, S)，其中 S 隔离了 A 和 B，成立

$$x_A \perp x_B \mid x_S$$

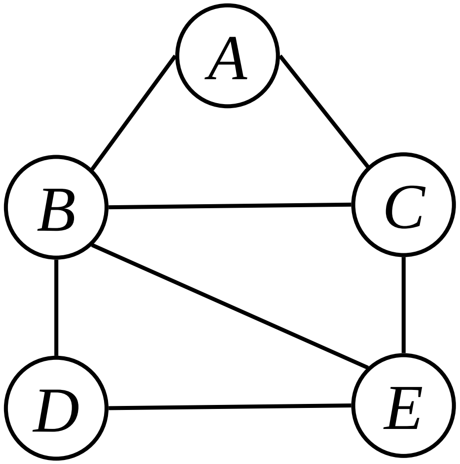
由依图g的全局Markov性诱导了一族分布: $\mathbb{M}_G(g)$



S 隔离了 A 和 B：从 A 到 B 的任一条路径都经过 S 中结点

Basic concepts – clique

clique 簇	C	称一个结点子集C为簇，如果C中结点在图中两两相邻
maximal clique 最大簇		不真包含于图中任何一个簇的簇



Definition - Factorization property (F)

称一个分布 $p(x_V)$ 服从依图g的分解性，如果对图中所有簇C，存在一个非负函数$\phi(x_C)$一称为势函数，使得

$$p(x_V)=\frac{1}{Z}\prod_{C\in\mathcal{C}}\phi(x_C) \qquad \text{或写成}\ p(x_V)\propto\prod_{C\in\mathcal{C}}\phi(x_C)$$

其中 Z 是归一化常数 $Z=\sum_{x_V}\prod_{C\in\mathcal{C}}\phi(x_C)$

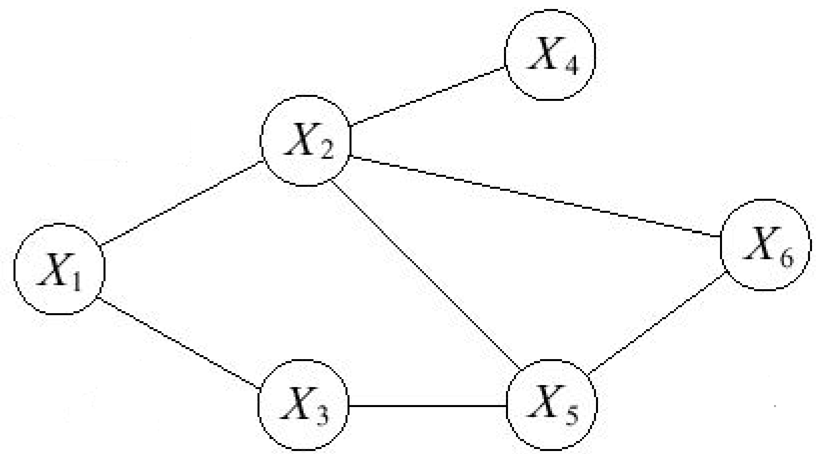
- $\mathcal{C}$ 是图中所有簇的集合
- 只要求簇上势函数的存在性。这些势函数不是唯一确定的
- 不失一般性，可定义在最大簇上

由依图g的分解性诱导了一族分布: $\mathbb{M}_F(g)$

The fundamental theorem of random fields (Hammersley & Clifford)

Theorem: $\mathbb{M}_G(g) = \mathbb{M}_F(g)$

UGM Example



$$p(x_1,x_2,x_3,x_4,x_5,x_6) \propto \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\phi(x_5)\phi(x_6)$$

=1

$$\times\, \phi(x_1,x_2)\phi(x_1,x_3)\phi(x_2,x_4)\phi(x_3,x_5)\phi(x_2,x_5)\phi(x_2,x_6)\phi(x_5,x_6)$$

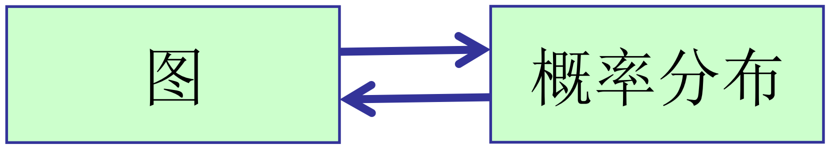
$$\times\, \phi(x_2,x_5,x_6)$$

$$\Updownarrow$$

$$p(x_1,x_2,x_3,x_4,x_5,x_6) \propto \psi(x_1,x_2)\psi(x_1,x_3)\psi(x_2,x_4)\psi(x_3,x_5)\psi(x_2,x_5,x_6)$$

UGM的语义

一个图表示了怎样的概率分布



一个概率分布如何表示成一个图

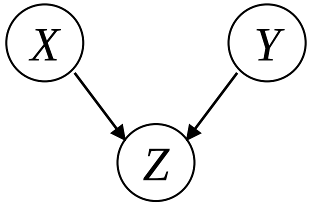
- 一个UGM表示一个怎样的概率分布（by definition）
  - 满足 分解性
  - 满足 Markov性
- 一个概率分布如何表示成一个UGM
  - 从条件独立性的包含关系上，寻找尽可能紧凑的I-map

P-map的存在性

- 不是每一个分布都有 完美无向图表示。

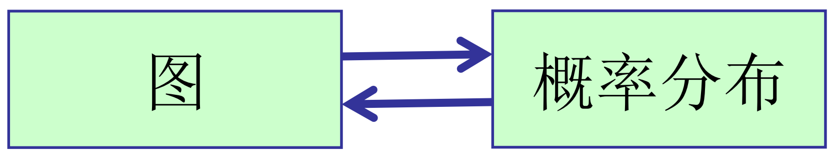
例子（KF p.122 example 4.8）

- 考虑三个随机变量X, Y, Z 组成的V连接的贝叶斯网络所代表的分布 p, 试寻找一个无向图，作为分布 p 的尽可能紧凑的I-map



UGM的语义

一个图表示了怎样的概率分布



一个概率分布如何表示成一个图

一个UGM表示一个怎样的概率分布（by definition）

- 满足 分解性
- 满足 Markov性

一个概率分布如何表示成一个UGM

- 从条件独立性的包含关系上，寻找尽可能紧凑的I-map

UGM适用于随机因素间彼此影响，通过局部函数的倾向性取值来表达的建模问题

- 局部函数$\phi(x_C)$ 体现 局部变量x_C取值的随机规律

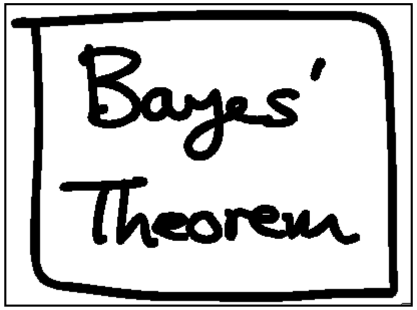
Image de-noising (PRML p387)

Given the noisy image, to recover the original clean image

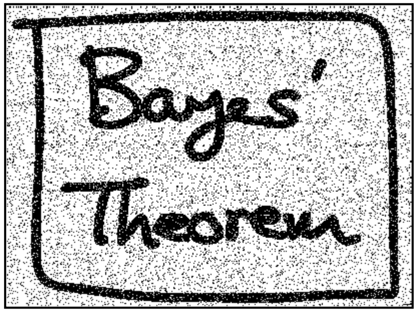
- x_i : original clean image pixel $\in \{-1, 1\}$, $i = 1, \ldots, N^2$
- y_i : observed noisy image pixel $\in \{-1, 1\}$

$$\hat{x}_{1:N^2} = \arg\max_{x_{1:N^2}} p\left(x_{1:N^2} \mid y_{1:N^2}\right)$$

$$p\left(x_{1:N^2}, y_{1:N^2}\right) = ?$$



original clean x

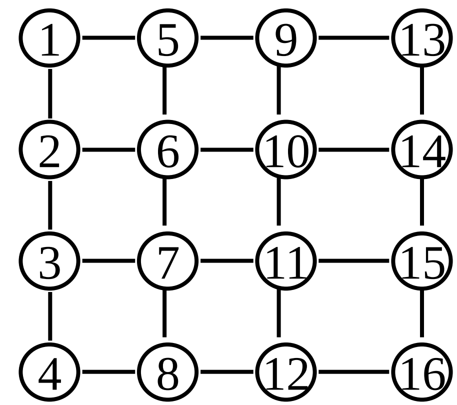


noisy observation y

Ising model

- Consider a lattice of binary RV's, $x_i \in \{-1,1\}$

$$p\left(x_{1:N^2}\right) \propto \exp\left\{\sum_{i-j}\phi\left(x_i, x_j\right)\right\} = \exp\left\{\beta\sum_{i-j} x_i x_j\right\} \quad \beta > 0$$



  - β表示 相邻变量取相同值 的可能性大小.
  - β 取不同值时Ising model的随机采样结果:

$\beta = ?$

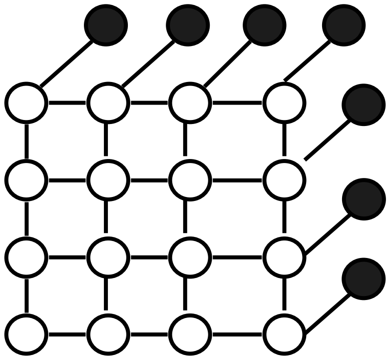
$\beta = 1$

$\beta = ?$

Image de-noising (PRML p387)

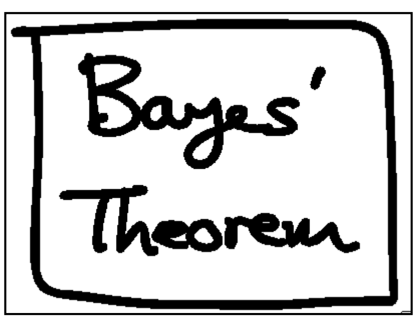
概率模型

- 相邻象素 x_i, x_j 取值的随机规律
- y_i 与 x_i 取值的随机规律

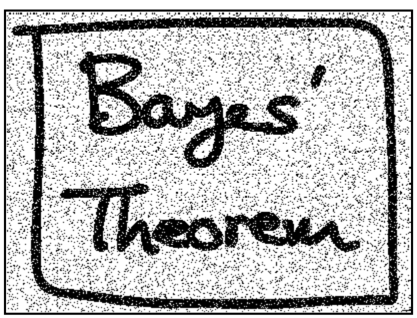


$$p(x,y) \propto \exp\left\{\beta \sum_{i \sim j} x_i x_j + \gamma \sum_i x_i y_i\right\} \quad \beta > 0, \gamma > 0$$

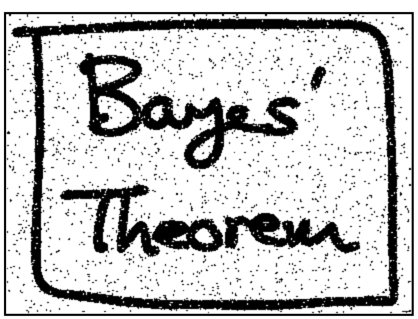
Compute $\hat{x} = \arg\max_x p(x \mid y)$?



original clean x



noisy observation y



estimate $\hat{x}$

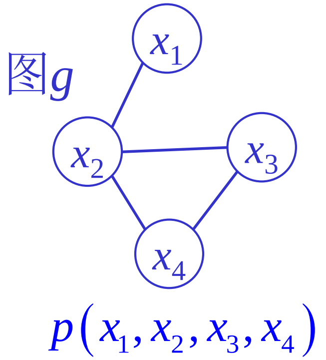
Graph semantics

无向图

Definition—无向图的F性质

- 称一个多变量的分布 $p(x_V)$ 服从依图g的分解性，如果对图中所有簇C，存在一个非负函数$\phi(x_C)$—称为势函数，使得

$$p\left(x_V\right)=\frac{1}{Z}\prod_{C\in\mathbf{C}}\phi\left(x_C\right)$$



或写成 $p\left(x_V\right)\propto\prod_{C\in\mathbf{C}}\phi\left(x_C\right)$

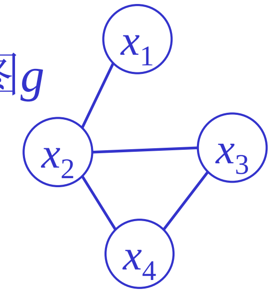
其中 Z 是归一化常数 $Z=\sum_{x}\prod_{C\in\mathbf{C}}\phi\left(x_C\right)$

- **C** 是图中所有簇的集合
- 只要求簇上势函数的存在性。这些势函数不是唯一确定的
- 不失一般性，可定义在最大簇上

Definition—无向图的P,L,G性质

- 称一个多变量的分布 $p(x_V)$ 服从

图g



$p(x_1, x_2, x_3, x_4)$

(P) 依图g的点对Markov性（pairwise Markov property），

如果对图g中任意不相邻的结点 α, β，成立

$$\alpha \perp \beta \mid V \setminus \{\alpha, \beta\}$$

(L) 依图g的局部Markov性（local Markov property），

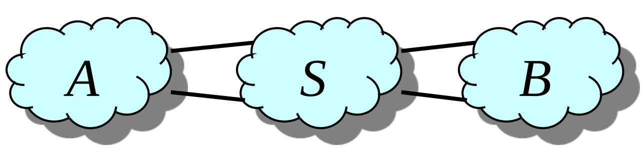
如果对图g中任意结点 α，成立

$$\alpha \perp V \setminus cl(\alpha) \mid bd(\alpha)$$

(G) 依图g的全局Markov性（global Markov property），

如果对图g中任意互不相交的三个结点子集 (A, B, S)，其中 S 隔离了 A 和 B，成立

$$A \perp B \mid S$$



Properties on undirected graphs

- For undirected graph g, (F)≡(G)≡(L)≡(P), if the probability distribution $p(x_V)$ is positive.

- The set of probability distributions is denoted by $\mathbb{M}(g)$.

$$(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)$$

If $p(X_V)$ is positive

Example 3.10 in Lauritzen's book p.37

P5 fail example

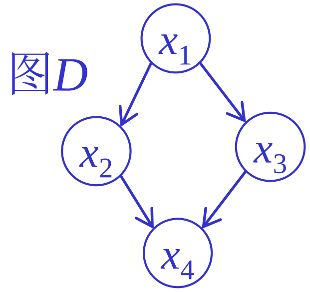
Graph semantics

有向无环图

Definition－DAG的(DF)性质

- 称一个多变量的分布 $p(x_V)$ 服从 依图D的有向分解性，如果

$$p(x_V)=\prod_{v\in V}p\left(x_v \mid x_{pa(v)}\right)$$



$p(x_1,x_2,x_3,x_4)$

Definition－DAG的(DO)(DL)性质

- 称一个多变量的分布 $p(x_V)$ 服从

(DO) 依图D的有向有序Markov性（directed ordered Markov property ），

如果在图D的某个良序下，成立

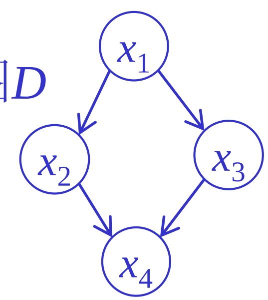
$$v \perp pr(v) \setminus pa(v) \mid pa(v)$$

(DL) 依图D的有向局部Markov性（ directed local Markov property），

如果对图D中任意结点 v，成立

$$v \perp nd(v) \setminus pa(v) \mid pa(v)$$

图D

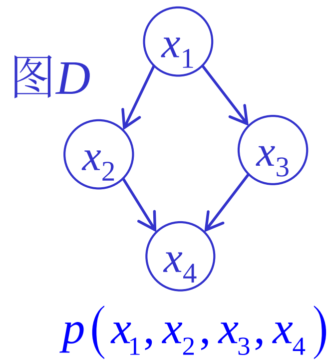


$p(x_1, x_2, x_3, x_4)$

Definition—DAG的(DG)性质

- 称一个多变量的分布 $p(x_V)$ 服从依图D的有向全局Markov性（directed global Markov property ），如果对图D中任意互不相交的三个结点子集 (A, B, S)，其中 S 有向隔离了 A 和 B，成立

$$A \perp B \mid S$$



Definition: d-separation

- 称结点集A和结点集B被结点集S有向隔离(d-separated)，如果从A到B的**所有**迹(all trails)都被S有向隔断(d-blocked)

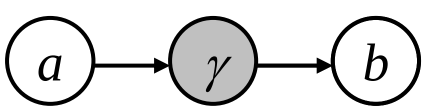
- 称从结点a到结点b的一条迹π 被 结点集 S 有向隔断

  如果 下列条件之一成立，

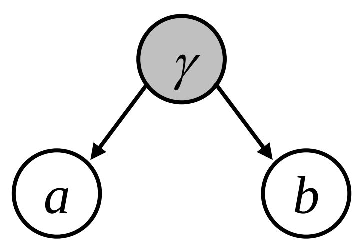
  ① 迹π 上**存在**一个结点 $\gamma \in \pi$， $\gamma \in S$，并且迹 π 在结点γ 处的箭头不是V连接

  ② 迹π 上**存在**一个结点 $\gamma \in \pi$， γ 与它的所有后代结点均 $\notin S$，并且迹 π 在结点γ 处的箭头是V连接

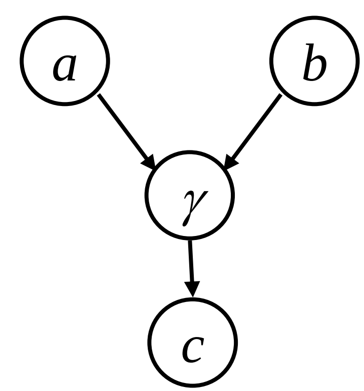
$S = \{\gamma\}$



$S = \{\gamma\}$

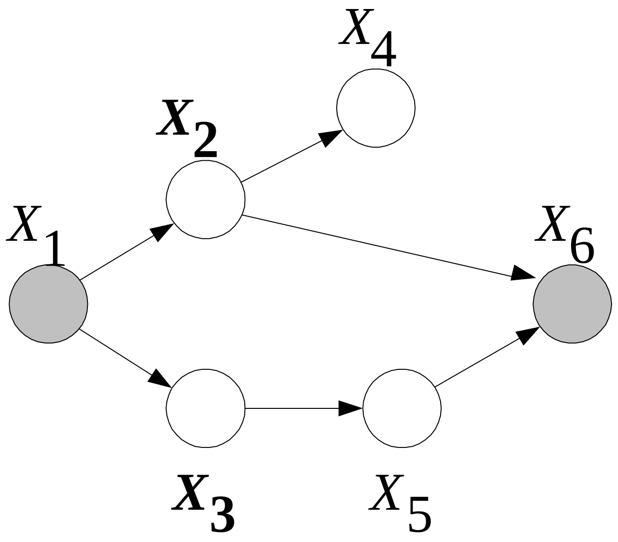


$S = \varnothing$



Properties on DAG

$$(DF) \Leftrightarrow (DG) \Leftrightarrow (DL) \Leftrightarrow (DO)$$



S

$$X_2 \perp X_3 \mid \{X_1, X_6\} \ ?$$

① 迹π上存在一个结点 $\gamma \in \pi$， $\gamma \in S$，并且迹 π 在结点γ处的箭头不是V连接

② 迹π上存在一个结点 $\gamma \in \pi$， γ 与它的所有后代结点均 $\notin S$，并且迹 π 在结点γ处的箭头是V连接

课程章节

- 第一章 引言（1）
- 第二章 图模型的表示理论（2）
  - **Semantics (DGM, UGM)**
  - HMM, CRF
- 第三章 图模型的推理理论（6）
  - 精确推理：variable-elimination，cluster-tree，triangulate
  - 连续变量：Kalman
  - 采样近似：sampling
  - 变分近似：variational
- 第四章 图模型的学习理论（3）
  - 参数学习：maxlikelihoodEstimate，UGMLearning, BayesEstimate
  - 结构学习：StructureLearning
- 第五章 一个综合例子（1）