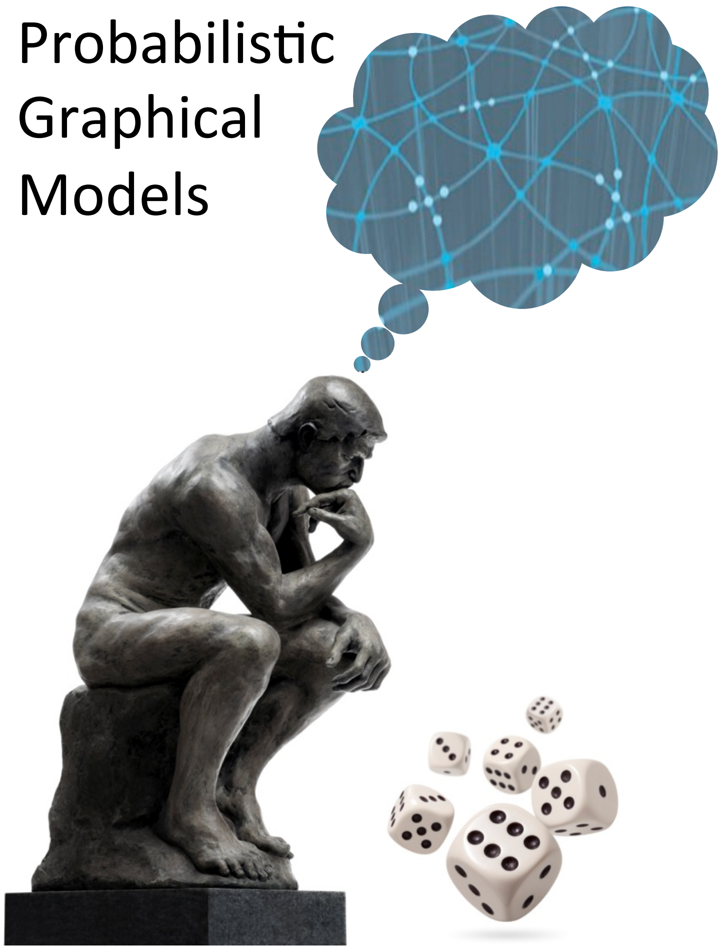
Probabilistic
Graphical
Models



Representation

Independencies

Bayesian Networks

因子化

Independence & Factorization

$P(X,Y) = P(X)\,P(Y)$ X,Y independent

$P(\mathbf{X},\mathbf{Y},\mathbf{Z}) \propto \phi_1(\mathbf{X},\mathbf{Z})\,\phi_2(\mathbf{Y},\mathbf{Z})$ $(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$

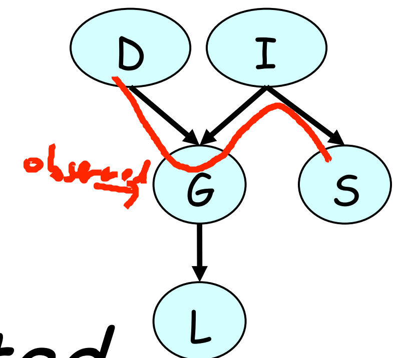
- Factorization of a distribution P implies independencies that hold in P
- If P factorizes over G, can we read these independencies from the structure of G?

Daphne Koller

有向分离 (Directed Separation)

Flow of influence & d-separation

有向分离，用于判断X,Y是否条件独立
有向分离可以直接从图上看出来，这样就能简化计算



Definition: **X** and **Y** are *d-separated* in G given **Z** if there is no active trail in G between **X** and **Y** given **Z**

Notation: d-sep$_G$(**X**, **Y** | **Z**)

Factorization ⇨ Independence: BNs

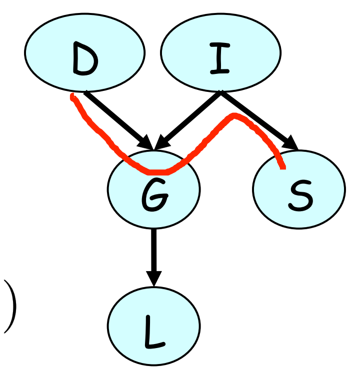
Theorem: If P factorizes over G, and d-sep$_G$(X, Y | Z) then P satisfies (X ⊥ Y | Z)

看懂这个推理

chain rule

$$\underline{P(D,I,G,S,L)} = \underbrace{P(D)P(I)P(G \mid D,I)P(S \mid I)P(L \mid G)}$$

$P \models D \perp S$



$$\underline{P(D,S)} = \sum_{G,L,I} P(D)P(I)P(G \mid D,I)P(S \mid I)\,P(L \mid G)$$

$$= \sum_{I} P(D)P(I)P(S \mid I) \underbrace{\sum_{G} (\overset{1}{\cancel{P(G \mid D,I)}} \underbrace{\sum_{L} P(\overset{1}{\cancel{L}} \mid G)})}$$

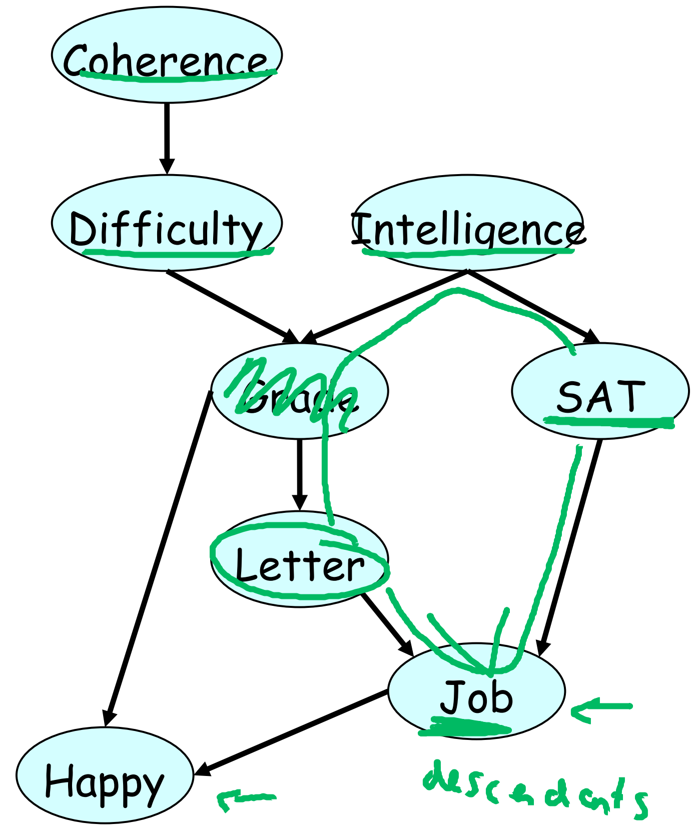
$$= \underset{\phi_1(D)}{P(D)}(\overbrace{\sum_{I} P(I)P(S \mid I)}^{\phi_2(S)})$$

Daphne Koller

Any node is d-separated from its non-descendants given its parents

Grade

If P factorizes over G, then in P, any variable is independent of its non-descendants given its parents



Daphne Koller

I-maps

- d-separation in $G \Rightarrow$ P satisfies corresponding independence statement

$$I(G) = \{(X \perp Y \mid Z) \ : \ \text{d-sep}_G(X, Y \mid Z)\}$$

- Definition: If P satisfies $I(G)$, we say that G is an I-map (independency map) of P

Daphne Koller

I-maps

$P_1 \models D \perp I$

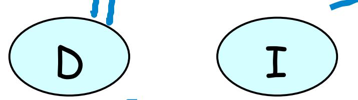
I	D	Prob
i^0	d^0	0.42
i^0	d^1	0.18
i^1	d^0	0.28
i^1	d^1	0.12

$P_2 \not\models D \perp I$

I	D	Prob.
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

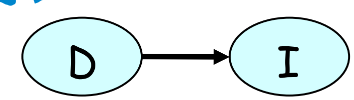
I-map

G_1



$I(G_1) = \{D \perp I\}$

G_2



$I(G_2) = \emptyset$

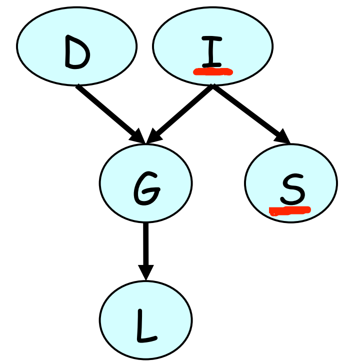
Daphne Koller

Factorization ⇨ Independence: BNs

Theorem: If P factorizes over G, then G is an I-map for P

Can read from G independencies in P
regardless of parameters

Independence ⇨ Factorization



Theorem: If G is an I-map for P, then P factorizes over G

$I \perp D$

chain rule for probabilities

$$P(D,I,G,S,L) = \overbrace{P(D)P(I \mid \cancel{D})}^{P(I,D)}P(G \mid D,I)P(S \mid \cancel{D},I,\cancel{G})P(L \mid \cancel{D},\cancel{I},G,\cancel{S})$$

$$P(D,I,G,S,L) = P(D)P(I)P(G \mid D,I)P(S \mid I)P(L \mid G)$$

Daphne Koller

Summary

Two equivalent views of graph structure:

- Factorization: G allows P to be represented
- I-map: Independencies encoded by G hold in P

If P factorizes over a graph G, we can read from the graph independencies that must hold in P (an independency map)