I-Maps: Graphs and Distributions

Sargur Srihari
srihari@cedar.buffalo.edu

Topics

- I-Maps
- I-Map to Factorization
- Factorization to I-Map
- Perfect Map

Graphs and Distributions

- Relating two concepts:
  - Independencies in distributions
  - Independencies in graphs
- I-Map is a relationship between the two

Independencies in a Distribution

- Let P be a distribution over X
- $I(P)$ is set of conditional independence assertions of the form $(X \perp Y | Z)$ that hold in P

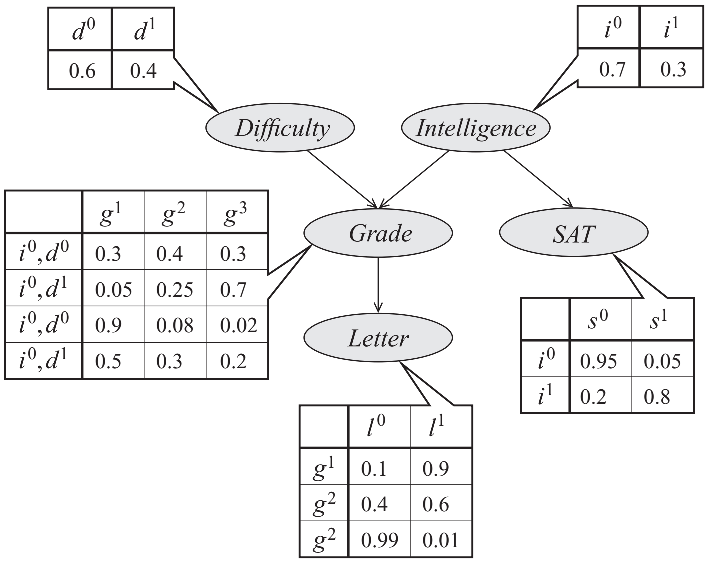
X	Y	$P(X,Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

X and Y are independent in P, e.g.,

$P(x^1)=0.48+0.12=0.6$
$P(y^1)=0.32+0.48=0.8$
$P(x^1,y^1)=0.48=0.6 \times 0.8$

Thus $(X \perp Y | \phi) \in I(P)$

Independencies in a Graph



- Graph G with CPDs is equivalent to a set of independence assertions

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

- Local Conditional Independence Assertions (starting from leaf nodes):

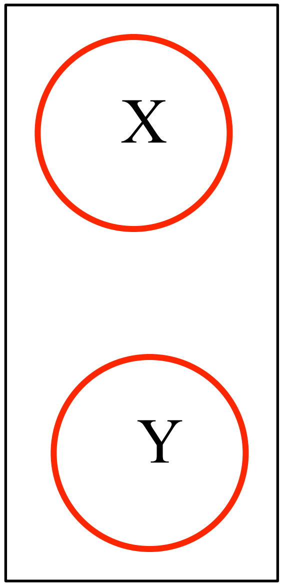
$I(G) = \{(L \perp I,D,S \mid G),$	L is conditionally independent of all other nodes given parent G
$(S \perp D,G,L \mid I),$	S is conditionally independent of all other nodes given parent I
$(G \perp S \mid D,I),$	Even given parents, G is NOT independent of descendant L
$(\mathrm{I} \perp D \mid \phi),$	Nodes with no parents are marginally independent
$(D \perp I,S \mid \phi)\}$	D is independent of non-descendants I and S

- Parents of a variable shield it from probabilistic influence
    - Once value of parents known, no influence of ancestors
- Information about descendants can change beliefs about a node

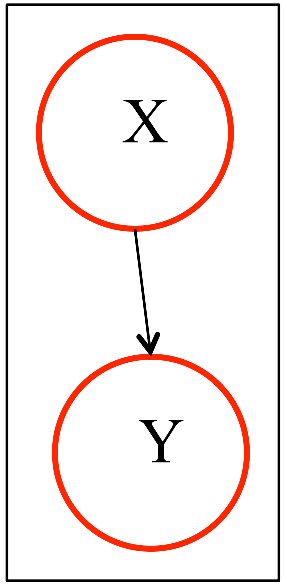
I-MAP

- Let G be a graph associated with a set of independencies $I(G)$
- Let P be a probability distribution with a set of independencies $I(P)$
- Then G is an I-map of I if $I(G) \subseteq I(P)$
- From direction of inclusion
  - distribution can have more independencies than the graph
  - Graph does not mislead in independencies existing in P

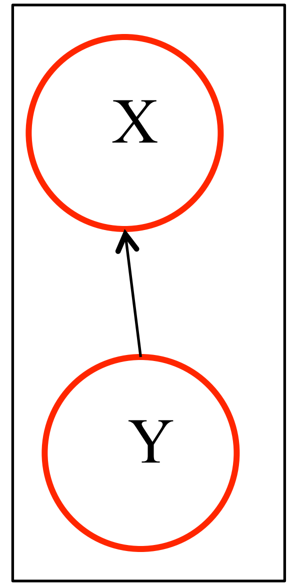
Example of I-MAP



G_0 encodes $X \perp Y$ or $I(G_0)=\{X \perp Y\}$



G_1 encodes no Independence or $I(G_1)=\{\Phi\}$



G_2 encodes no Independence $I(G_2)=\{\Phi\}$

X	Y	$P(X,Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

X and Y are independent in P, e.g.,

G_0 is an I-map of P
G_1 is an I-map of P
G_2 is an I-map of P

X	Y	$P(X,Y)$
x^0	y^0	0.4
x^0	y^1	0.3
x^1	y^0	0.2
x^1	y^1	0.1

X and Y are not independent in P
Thus $(X \perp Y) \notin I(P)$

G_0 is not an I-map of P
G_1 is an I-map of P
G_2 is an I-map of P

If G is an I-map of P then it captures some of the independences, not all

I-map to Factorization

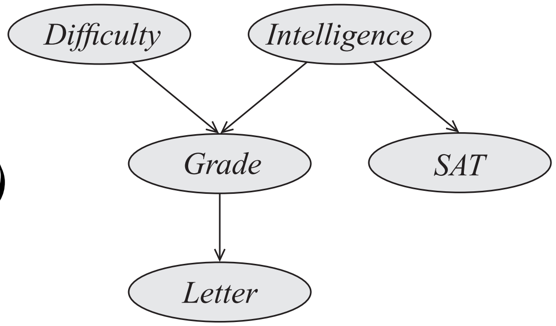
- A Bayesian network G encodes a set of conditional independence assumptions $I(G)$
- Every distribution P for which G is an I-map should satisfy these assumptions
  - Every element of $I(G)$ should be in $I(P)$
- This is the key property to allowing a compact representation

I-map to Factorization

- From chain rule of probability

$$P(I,D,G,L,S)=P(I)P(D|I)P(G|I,D)P(L|I,D,G)P(S|I,D,G,L)$$

  - Relies on no assumptions
  - Also not very helpful
    - Last factor requires evaluation of 24 conditional probabilities
- Apply conditional independence assumptions induced from the graph



$$D \perp I \in I(P) \text{ therefore } P(D|I)=P(D)$$

$$(L \perp I,D) \in I(P) \text{ therefore } P(L|I,D,G)=P(L|G)$$

  - Thus we get

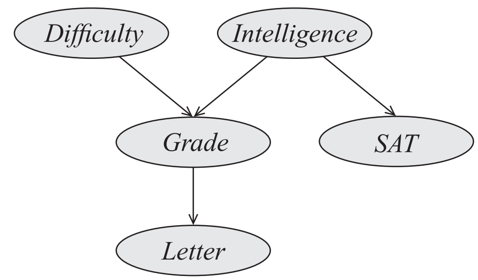
$$P(D,I,G,S,L) = P(D)P(I)P(G \mid D,I)P(S \mid I)P(L \mid G)$$

    - Which is a factorization into local probability models
  - Thus we can go from graphs to factorization of P

Factorization to I-map

- We have seen that we can go from the independences encoded in G, i.e., $I(G)$, to Factorization of P
- Conversely, Factorization according to G implies associated conditional independences
  - If P factorizes according to G then G is an I-map for P
  - Need to show that if P factorizes according to G then $I(G)$ holds in P
  - Proof by example

Example that independences in $\mathcal{G}$ hold in P



- P is defined by set of CPDs
- Consider independences for S in G, i.e.,

$$P(S \perp D,G,L|I)$$

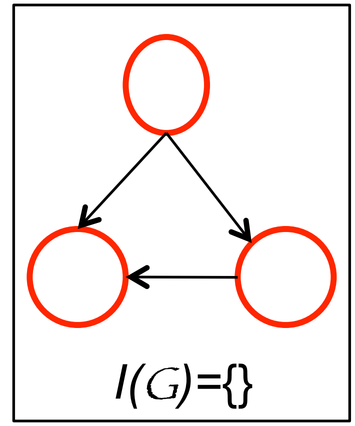
- Starting from factorization induced by graph

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

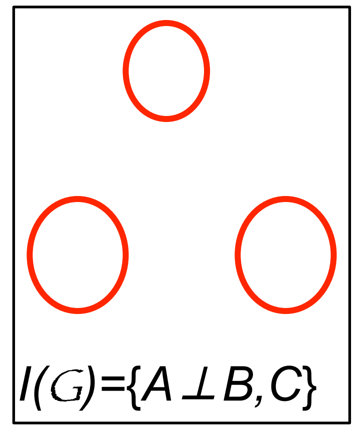
- Can show that $P(S|I,D,G,L)=P(S|I)$
- Which is what we had assumed for P

Perfect Map

- I-map
  - All independencies in $I(G)$ present in $I(P)$
  - Trivial case: all nodes interconnected



- D-Map
  - All independencies in $I(P)$ present in $I(G)$
  - Trivial case: all nodes disconnected



- Perfect map
  - Both an I-map and a *D*-map
  - Interestingly not all distributions P over a given set of variables can be represented as a perfect map
    - Venn Diagram where *D* is set of distributions that can be represented as a perfect map