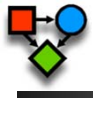


Readings: K&F 3.1, 3.2, 3.3, 3.4.1



# Bayesian Network Representation

---

Lecture 2 – Mar 30, 2011  
CSE 515, Statistical Methods, Spring 2011

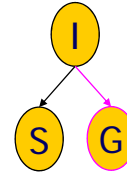
Instructor: Su-In Lee  
University of Washington, Seattle

## Last time & today

- Last time
  - Probability theory
  - Conditional independence
  - Conditional parameterization
  
- Today
  - Naïve Bayes model
  - Definition of the Bayesian network (BN)
  - Independence properties encoded in BN graphs
  - From distributions to BN graphs

## Conditional parameterization

- S = SAT score,  $\text{Val}(S) = \{s^0, s^1\}$
  - I = Intelligence,  $\text{Val}(I) = \{i^0, i^1\}$
  - G = Grade,  $\text{Val}(G) = \{g^0, g^1, g^2\}$
  - Assume that G and S are independent given I
- $P(I, S, G) =$



### Joint parameterization

P(I,S)G		P(I,S,G)	
I	S	P(I,S)	P(I,S,G)
i <sup>0</sup>	s <sup>0</sup>	0.665	0.425
i <sup>0</sup>	s <sup>1</sup>	0.035	0.125
i <sup>0</sup>	s <sup>0</sup>	0.06	
i <sup>1</sup>	s <sup>0</sup>	0.24	0.01
:	:	:	:

3 parameters  
11 parameters

### Conditional parameterization

P(I)		P(S I)		P(G I)		
I		S		I	G	
		s <sup>0</sup>	s <sup>1</sup>		g <sup>0</sup>	g <sup>1</sup> g <sup>2</sup>
i <sup>0</sup>	i <sup>1</sup>			i <sup>0</sup>	0.75	0.05 0.2
0.7	0.3	i <sup>0</sup>	0.95 0.05	i <sup>1</sup>	0.2	0.3 0.5
		i <sup>1</sup>	0.2 0.8			

3 parameters 7 parameters

## Naïve Bayes model

- Class variable C,  $\text{Val}(C) = \{c_1, \dots, c_k\}$
- Evidence variables  $X_1, \dots, X_n$
- Naïve Bayes assumption: evidence variables are conditionally independent given C

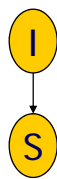


$$P(C, X_1, \dots, X_n) =$$

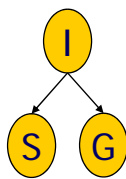
- Applications in medical diagnosis, text classification
- Used as a classifier:
  - Given  $\{x_1, \dots, x_n\}$  on evidence variables  $X_1, \dots, X_n$ , predict the value on C :
 
$$\frac{P(C = c_1 | x_1, \dots, x_n)}{P(C = c_2 | x_1, \dots, x_n)}$$
- Problem: Double counting correlated evidence

## Bayesian network (informal)

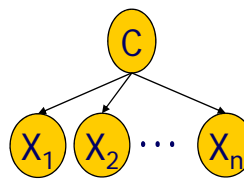
- Directed acyclic graph (DAG)  $G$ 
  - Nodes represent random variables
  - Edges represent direct influences between random variables
- Local probability models (conditional parameterization)
  - Conditional probability distributions (CPDs)
- Here are the networks we have been discussing so far...



Example 1



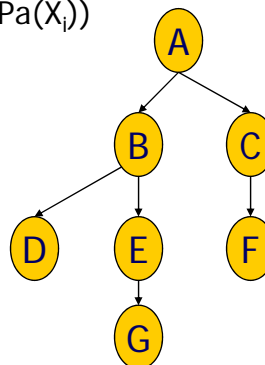
Example 2



Naïve Bayes

## Bayesian network structure

- Directed acyclic graph (DAG)  $G$ 
  - Nodes  $X_1, \dots, X_n$  represent random variables
- $G$  encodes the following set of independence assumptions (called, **local independencies**)
  - $X_i$  is independent of its non-descendants given its parents
  - Formally:  $(X_i \perp \text{NonDesc}(X_i) \mid \text{Pa}(X_i))$
  - Denoted by  $I_L(G)$



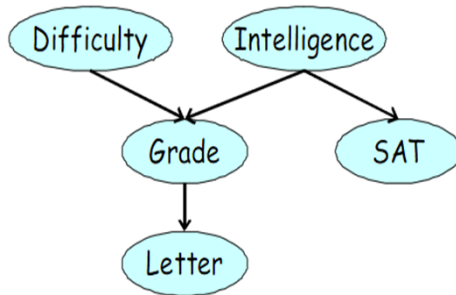
$$E \perp \{A, C, D, F\} \mid B$$



## The *Student* example

- Course difficulty (**D**), Val(D) = {easy, hard}
- Intelligence (**I**), Val(I) = {high, low}
- Grade (**G**), Val(G) = {A, B, C}
- Quality of the rec. letter (**L**), Val(L) = {strong, weak}
- SAT (**S**), Val(S) = {high, low}

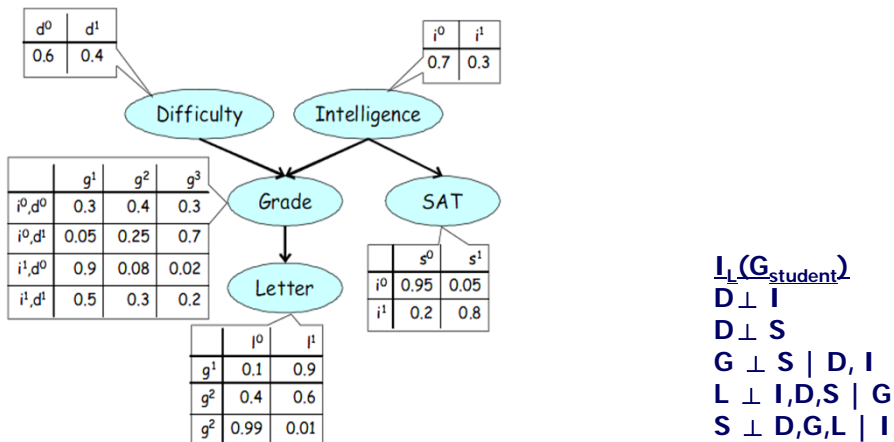
**Graph  $G_{\text{student}}$**



Local independencies  $I_L(G_{\text{student}})$

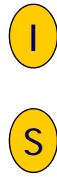
## The *Student* Bayesian network

- Joint distribution
  - $P(I, D, G, S, L) =$



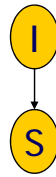
## Independency mappings (I-Maps)

- Let  $P$  be a distribution over  $X$
- Let  $I(P)$  be the independencies  $(X \perp Y \mid Z)$  in  $P$
- A Bayesian network structure  $G$  is an I-map (independency mapping) for  $P$ , if  $I_L(G) \subseteq I(P)$



I	S	P(I,S)
$i^0$	$s^0$	0.25
$i^0$	$s^1$	0.25
$i^1$	$s^0$	0.25
$i^1$	$s^1$	0.25

$$I_L(G) = \{I \perp S\} \quad I(P) = \{I \perp S\}$$



I	S	P(I,S)
$i^0$	$s^0$	0.4
$i^0$	$s^1$	0.3
$i^1$	$s^0$	0.2
$i^1$	$s^1$	0.1

$$I_L(G) = \emptyset \quad I(P) = \emptyset$$

## Factorization theorem

***“P factorizes over G”***

- $G$  is an I-Map of  $P \rightarrow P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa(X_i))$ 
  - The conditional independencies encoded in  $G$  imply factorization according to  $G$ .
- $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid Pa(X_i)) \rightarrow G$  is an I-Map of  $P$ 
  - Factorization according to  $G$  implies the associated conditional independencies.

## Factorization theorem

- If  $G$  is an I-Map of  $P$ , then  $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i))$

**Proof:**

- wlog.  $X_1, \dots, X_n$  is an ordering consistent with  $G$
- By chain rule:  $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1})$
- From assumption:  $Pa(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$   
 $\{X_1, \dots, X_{i-1}\} - Pa(X_i) \subseteq NonDesc(X_i)$
- Since  $G$  is an I-Map  $\rightarrow (X_i; NonDesc(X_i) | Pa(X_i)) \in I(P)$



$$P(X_i | X_1, \dots, X_{i-1}) = P(X_i | Pa(X_i))$$

11

## Factorization implies I-Map

- $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \rightarrow G$  is an I-Map of  $P$

**Proof:**

- Need to show  $(X_i; NonDesc(X_i) | Pa(X_i)) \in I(P)$  or that  $P(X_i | NonDesc(X_i)) = P(X_i | Pa(X_i))$
- wlog.  $X_1, \dots, X_n$  is an ordering consistent with  $G$

$$\begin{aligned} P(X_i | NonDesc(X_i)) &= \frac{P(X_i, NonDesc(X_i))}{P(NonDesc(X_i))} \\ &= \frac{\prod_{k=1}^i P(X_k | Pa(X_k))}{\prod_{k=1}^{i-1} P(X_k | Pa(X_k))} \\ &= P(X_i | Pa(X_i)) \end{aligned}$$

12

## Bayesian network definition

- A Bayesian network is a pair  $(G,P)$ 
  - $P$  factorizes over  $G$
  - $P$  is specified as set of CPDs associated with  $G$ 's nodes
- Parameters
  - Joint distribution:  $2^n$
  - Bayesian network (bounded in-degree  $k$ ):  $n2^k$

## Bayesian network design

- Variable considerations
  - Clarity test: can an omniscient being determine its value?
  - Hidden variables?
  - Irrelevant variables
- Structure considerations
  - Causal order of variables
  - Which independencies (approximately) hold?
- Probability considerations
  - Zero probabilities
  - Orders of magnitude
  - Relative values

## Independencies in a BN

- G encodes **local independencies**
  - $X_i$  is independent of its non-descendants given its parents
  - Formally:  $(X_i \perp \text{NonDesc}(X_i) \mid \text{Pa}(X_i))$

Does G encode other independence assumptions that hold in every distribution P that factorizes over G?



Devise a procedure to find all independencies in G

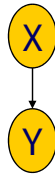
## d-Separation (directed separation)

- **Goal:** procedure that  $\text{d-sep}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z}, G)$ 
  - Return “true” iff  $\text{Ind}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$  follows from the local independencies in G,  $I_L(G)$ .
- **Strategy:** since influence must “flow” along paths in G, consider reasoning patterns between  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$ , in various structures in G
- **Active path:** creates dependencies between nodes
- **Inactive path:** cannot create dependencies



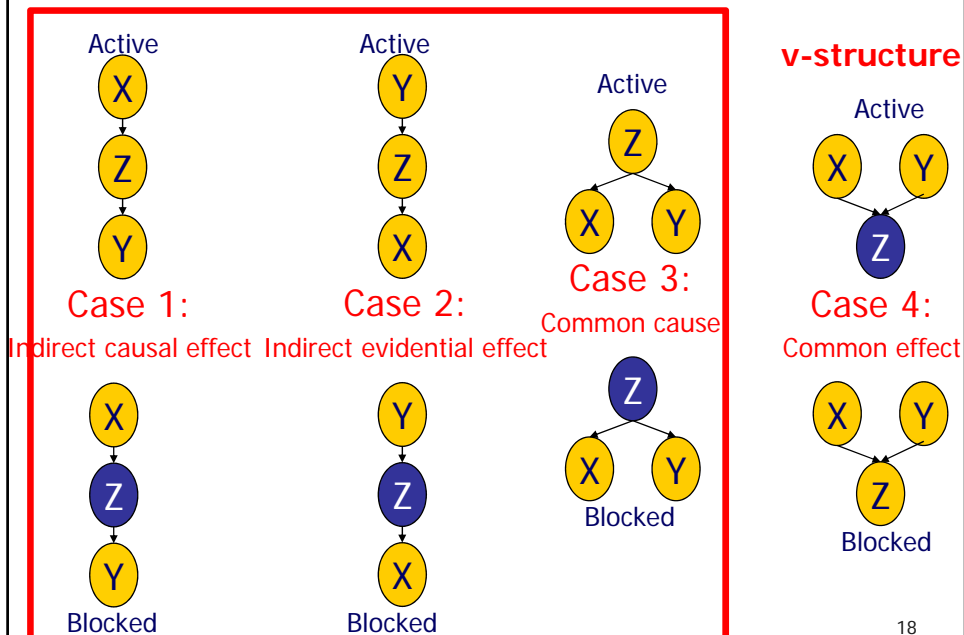
## Direct connection

- **X** and **Y** directly connected in  $G \rightarrow$  no  $Z$  exists for which  $\text{Ind}(X;Y | Z)$ 
  - Example: deterministic function



## Indirect connection

**X can influence Y via Z iff Z is not observed.**



## The general case

- Let  $G$  be a Bayesian network structure
- Let  $X_1 \leftrightarrow \dots \leftrightarrow X_n$  be a trail in  $G$
- Let  $\mathbf{E}$  be a subset of evidence nodes in  $G$



The trail  $X_1 \leftrightarrow \dots \leftrightarrow X_n$  is active given evidence  $\mathbf{E}$  if:

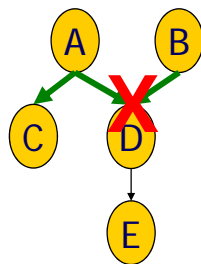
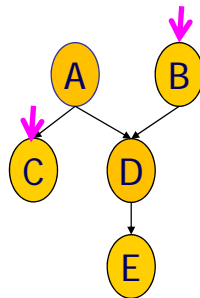
- ALL the three-node networks along the trail is active.
  - For every V-structure  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ ,  $X_i$  or one of its descendants is observed
  - No other nodes along the trail is in  $\mathbf{E}$

## d-Separation

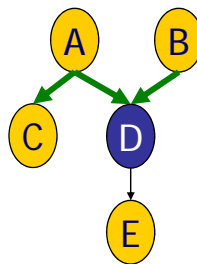
- $\mathbf{X}$  and  $\mathbf{Y}$  are **d-separated** in  $G$  given  $\mathbf{Z}$ , denoted  $d\text{-sep}_G(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$  if there is no active trail between any node  $X \in \mathbf{X}$  and any node  $Y \in \mathbf{Y}$  in  $G$
- $I(G) = \{(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}) : d\text{-sep}_G(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})\}$

## Examples

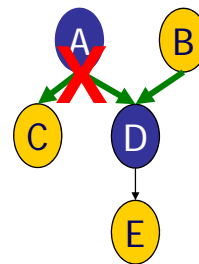
Are B and C d-separated?



d-sep(B,C)=yes



d-sep(B,C|D)=no



d-sep(B,C|A,D)=yes

## d-Separation: soundness

- Theorem:

- G is an I-map of P
- d-sep<sub>G</sub>(X;Y | Z) = yes



- P satisfies Ind(X;Y | Z)

- Defer proof

## d-Separation: completeness

### Theorem:

▪  $d\text{-sep}_G(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z}) = \text{no}$



There exists P such that

- G is an I-map of P
- P does not satisfy  $\text{Ind}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z})$

### Proof outline:

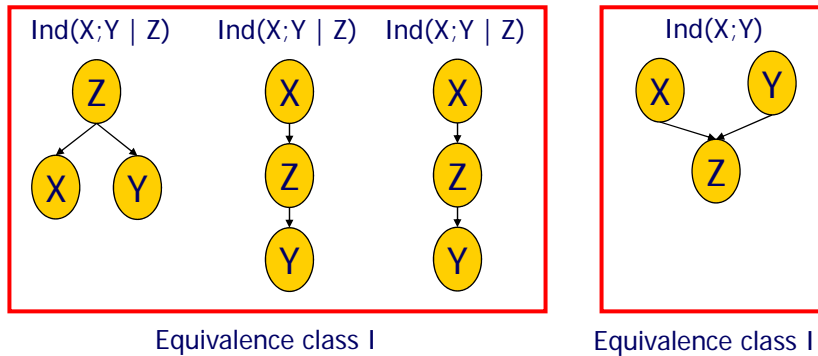
- Construct distribution P where independence does not hold
- Since there is no d-sep, there is an active path
- For each interaction in the path, correlate the variables through the distribution in the CPDs
- Set all other CPDs to uniform, ensuring that influence flows only in a single path and cannot be cancelled out
- Detailed distribution construction quite involved

## Algorithm for d-separation

- Goal: answer whether  $d\text{-sep}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z}, G)$ 
  - Enumerate all possible trails between X and Y? NO
- Algorithm:
  - Mark all nodes in  $\mathbf{Z}$  or that have descendants in  $\mathbf{Z}$
  - BFS traverse G from  $\mathbf{X}$
  - Stop traversal at blocked nodes:
    - Node that is in the middle of a v-structure and not in marked set
    - Not such a node but is in  $\mathbf{Z}$
  - If we reach any node in  $\mathbf{Y}$  then there is an active path and thus  $d\text{-sep}(\mathbf{X}; \mathbf{Y} \mid \mathbf{Z}, G)$  does not hold
- Theorem: algorithm returns all nodes reachable from  $\mathbf{X}$  via trails that are active in G

## I-equivalence between graphs

- $I(G)$  describe all conditional independencies in  $G$
- Different Bayesian networks can have same  $Ind$ .



Two BN graphs  $G_1$  and  $G_2$  are **I-equivalent** if  $I(G_1) = I(G_2)$

CSE 515 – Statistical Methods – Spring 2011

25

## I-equivalence between graphs

- If  $P$  factorizes over a graph in an I-equivalence class
  - $P$  factorizes over all other graphs in the same class
  - $P$  cannot distinguish one I-equivalent graph from another
- Implications for structure learning
  - We cannot find the “correct” structure from within the same equivalent class. -> will revisit later.
- Test for I-equivalence: d-separation

CSE 515 – Statistical Methods – Spring 2011

26

## Test for I-equivalence

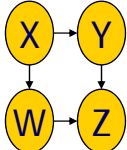
- **Necessary condition:** same graph skeleton
  - Otherwise, can find active path in one graph but not other
  - But, not sufficient: v-structures
- **Sufficient condition:** same skeleton and v-structures
  - But, not necessary: complete graphs (no independence)
- Define  $X \rightarrow Z \leftarrow Y$  as **immoral** if  $X, Y$  are not directly connected
  - **Necessary and Sufficient:** same skeleton and immoral set of v-structures

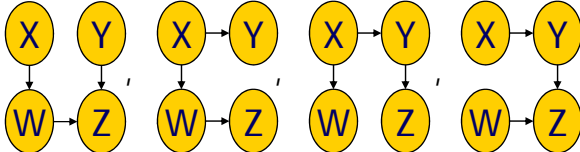
## Constructing graphs for P

- Can we construct a graph for a distribution P?
  - Any graph which is an I-map for P
  - But, this is not so useful: complete graphs
    - A DAG is complete if adding an edge creates cycles
    - Complete graphs imply no independence assumptions
    - Thus, they are I-maps of any distribution

## Minimal I-Maps

- A graph  $G$  is a minimal I-Map for  $P$  if:
  - $G$  is an I-map for  $P$
  - Removing any edge from  $G$  renders it not an I-map

- Example: if  is a minimal I-map for  $P$ ,

- Then:
- 
- is not I-maps.

## BayesNet definition revisited

- A Bayesian network is a pair  $(G,P)$ 
  - $P$  factorizes over  $G$
  - $P$  is specified as set of CPDs associated with  $G$ 's nodes
  - **Additional requirement:  $G$  is a minimal I-map for  $P$**

## Constructing minimal I-Maps

- Reverse factorization theorem
  - $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \rightarrow G$  is an I-Map of  $P$
- Algorithm for constructing a minimal I-Map
  - Fix an ordering of nodes  $X_1, \dots, X_n$
  - Select parents of  $X_i$  as minimal subset of  $X_1, \dots, X_{i-1}$ , such that  $\text{Ind}(X_i ; X_1, \dots, X_{i-1} - Pa(X_i) | Pa(X_i))$
- (Outline of) Proof of minimal I-map
  - I-map since the factorization above holds by construction
  - Minimal since by construction, removing one edge destroys the factorization

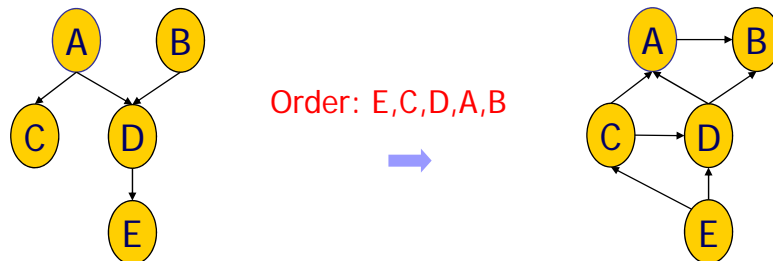
CSE 515 – Statistical Methods – Spring 2011

31

## Non-uniqueness of minimal I-Map

- Applying the same I-Map construction process with **different orders** can lead to different structures

Assume:  $I(G) = I(P)$



Different independence assumptions (different skeletons, e.g.,  $\text{Ind}(A;B)$  holds on left)

CSE 515 – Statistical Methods – Spring 2011

32

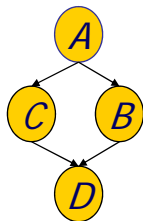


## Choosing order

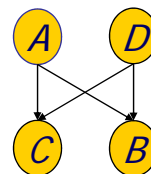
- Drastic effects on complexity of minimal I-Map graph
- Heuristic: use **causal** order

## Perfect maps

- G is a **perfect map (P-Map)** for P if  $I(P) = I(G)$
- Does every distribution have a P-Map?
  - No: independencies may be encoded in CPD  $\text{Ind}(X;Y|Z=1)$
  - No: some structures cannot be represented in a BN
    - Independencies in P:  $\text{Ind}(A;D | B,C)$ , and  $\text{Ind}(B;C | A,D)$



$\text{Ind}(B;C | A,D)$  does not hold



$\text{Ind}(A,D)$  also holds

## Finding a perfect map

- If P has a P-Map, can we find it?
  - Not uniquely, since I-equivalent graphs are indistinguishable
  - Thus, represent I-equivalent graphs and return it
- Recall I-Equivalence
  - **Necessary and Sufficient:** same skeleton and immoral set of v-structures
- Finding P-Maps
  - Step I: Find skeleton
  - Step II: Find immoral set of v-structures
  - Step III: Direct constrained edges

CSE 515 – Statistical Methods – Spring 2011

35

## Summary

- **Local independencies**  $I_L(G)$  – basic BN independencies
- **d-separation** – all independencies via graph structure
- G is an **I-Map** of P if and only if P factorizes over G
- **I-equivalence** – graphs with identical independencies
- **Minimal I-Map**
  - All distributions have I-Maps (sometimes more than one)
  - Minimal I-Map does not capture all independencies in P
- **Perfect map** – not every distribution P has one
  
- **Reading assignment:** K&F 3.1, 3.2, 3.3, 3.4
- **HW1 will be handed out next Monday!**

## Acknowledgement

- These lecture notes were generated based on the slides from Prof Eran Segal.