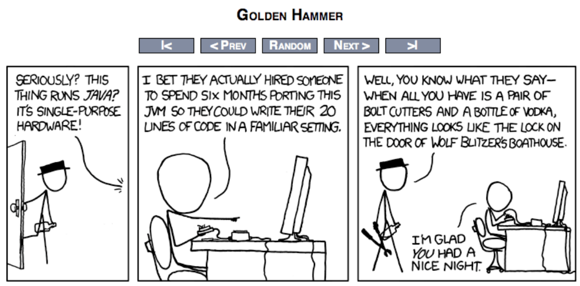


http://www.xkcd.com/801/

Bayes nets

CS151
David Kauchak
Fall 2010

Some material borrowed from:
Sara Owsley Sood and others

Independence

- Two variables are independent if knowing the values of one, does not give us information about the other
  - P(A,B) = P(A)P(B)
  - P(A|B) = P(A)
- Variables can also be independent only when they are conditioned on another variable
  - P(A,B|C) = P(A|C)P(B|C)
  - P(A|B,C) = P(A|C)
- Why do we care about variable independence?

Cavities

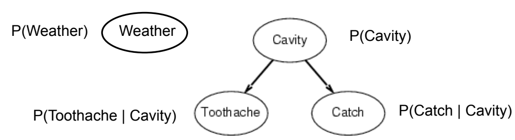
$$P(W,CY,T,CH) = P(W)P(CY)P(T \mid CY)P(CH \mid CY)$$

What independences are encoded
(both unconditional and conditional)?

Bayes nets

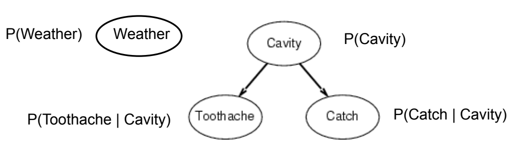
- Bayes nets are a way of representing joint distributions
  - Directed, acyclic graphs
  - Nodes represent random variables
  - Directed edges represent dependence
  - Associated with each node is a conditional probability distribution
    - P(X | parents(X))
  - They encode dependences/independences

Cavities



What independences are encoded?

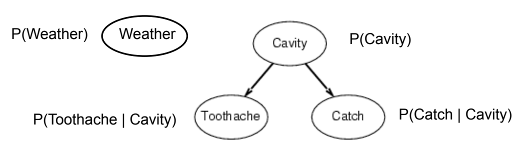
Cavities



Weather is independent of all variables
Toothache and Catch are conditionally independent GIVEN Cavity

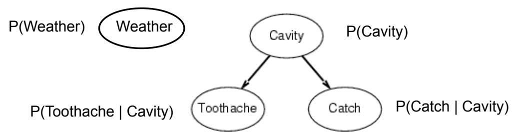
Does this help us in storing the distribution?

Why all the fuss about independences?



- Basic joint distribution
  - $2^4 = 16$ entries
- With independences?
  - 2 + 2 + 4 + 4 = 12
  - If we're sneaky: 1 + 1 + 2 + 2 = 6
  - Can be much more significant as number of variables increases!

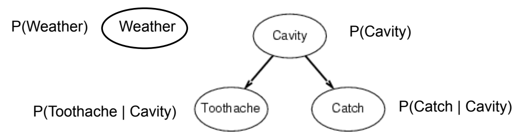
Cavities



$$P(W,T,CY,CH) = P(W)P(T,CY,CH \mid W)$$
$$= P(W)P(CY \mid W)P(T,CH \mid CY,W)$$
$$= P(W)P(CY \mid W)P(CH \mid CY,W)P(T \mid CH,CY,W)$$

Independences?

Cavities



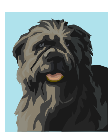
$$= P(W)P(CY)P(CH \mid CY)P(T \mid CY)$$

Graph allows us to figure out dependencies, and encodes the same information as the joint distribution.

Another Example

Question: Is the family next door out?

Variables that give information about this question:
- DO: is the dog outside?
- FO: is the family out (away from home)?
- LO: are the lights on?
- BP: does the dog have a bowel problem?
- HB: can you hear the dog bark?



Exploit Conditional Independence

Which variables are directly dependent?

Variables that give information about this question:
- DO: is the dog outside?
- FO: is the family out (away from home)?
- LO: are the lights on?
- BP: does the dog have a bowel problem?
- HB: can you hear the dog bark?

Are LO and DO independent?
What if you know that the family is away?

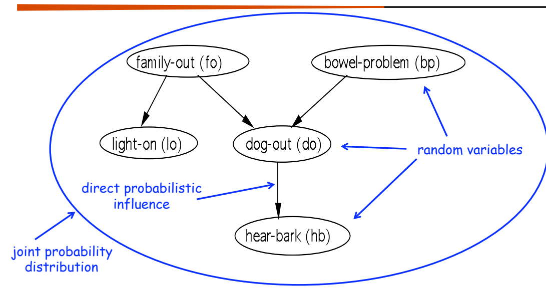
Are HB and FO independent?
What if you know that the dog is outside?

Some options

- lights (LO) depends on family out (FO)
- dog out (DO) depends on family out (FO)
- barking (HB) depends on dog out (DO)
- dog out (DO) depends on bowels (BP)

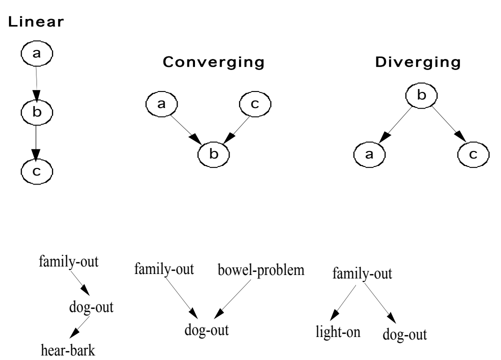
What would the network look like?

Bayesian Network Example

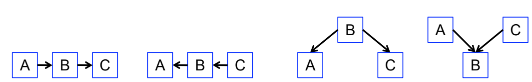


Graph structure represents direct influences between variables
(Can think of it as causality—but it doesn't have to be)

Three Types of Relationships



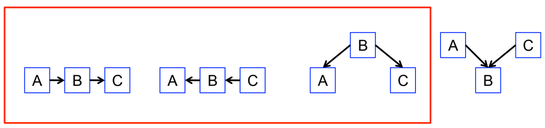
Relationships and independence



In which of these are A and C dependent/independent?

d-separation

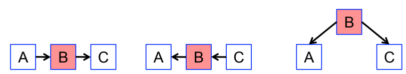
- "dependence" separation
- an *active* path in a BN is a path that carries information
  - If we know A does that give us information about C?
- Two variables are dependent if there is an active path between them



Which have active paths from A to B?

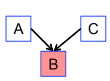
d-separation

- We can "block" an active path by conditioning on the internal node (B)
  - The dependence between A and C is through B
- A and C are independent given B



d-separation

- In this case, A and C are already independent

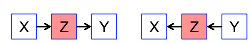


What happens when we condition on B?

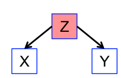
We make A and C dependent!
This is sometimes referred to as the "explaining away" phenomenon

d-separation

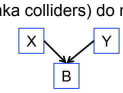
- *sets* of nodes **X** and **Y** are independent *given* a set of nodes **Z** if the sets are d-separated by **Z**
- **Z** d-separates **X** and **Y** if for all undirected paths from **X** to **Y**
  - the path contains a "chain" with a node from **Z** in the middle



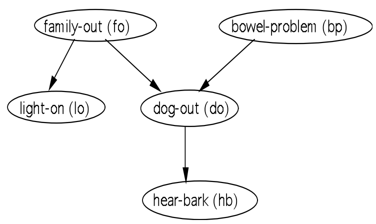
  - the path contains a "fork" with a node from **Z** in the middle



  - Any inverted "forks" (aka colliders) do not contain a node from **Z**

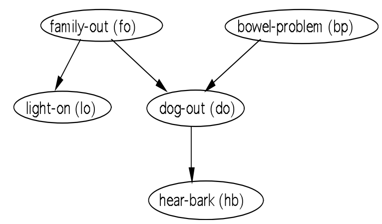


Independence



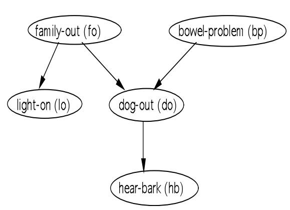
What unconditional independences does this graph encode, i.e. what nodes are d-separated using an empty set?

Independence



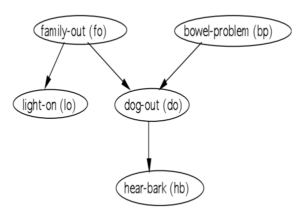
FO independent of BP
LO independent of BP

Conditional Independence



What conditional independences does this graph encode?

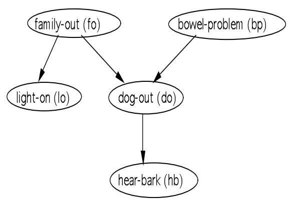
Conditional Independence



HB independent of FO, BP, LO given DO
DO independent of LO given FO
LO independent of HB, DO, BP given FO

Markov Blanket

- The Markov blanket of a node is:
  - the parents
  - the children
  - the parents of the children
- A node is independent of all other nodes, given its Markov blanket



How do these independences help?

Question: Is the family next door out?

Variables that give information about this question:
- DO: is the dog outside?
- FO: is the family out (away from home)?
- LO: are the lights on?
- BP: does the dog have a bowel problem?
- HB: can you hear the dog bark?

How many entries in the full joint distribution table?

How do these independences help?

Question: Is the family next door out?

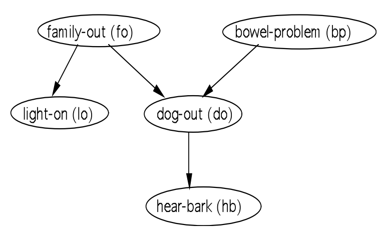
Variables that give information about this question:
- DO: is the dog outside?
- FO: is the family out (away from home)?
- LO: are the lights on?
- BP: does the dog have a bowel problem?
- HB: can you hear the dog bark?

32

Joint Probability Table

		hb				¬hb			
		do		¬do		do		¬do	
		bp	¬bp	bp	¬bp	bp	¬bp	bp	¬ bp
fo	lo								
	¬ lo								
¬fo	lo								
	¬ lo								

How many in the Bayes Net?



How do BNs help?

Joint Probability Table

		hb				¬hb			
		do		¬do		do		¬do	
		bp	¬bp	bp	¬bp	bp	¬bp	bp	¬ bp
fo	lo								
	¬ lo								
¬fo	lo								
	¬ lo								

vs

Conditional Probability Tables

P(HB|DO)

	hb	¬hb
do		
¬do		

P(FO)

fo	¬fo

P(BP)

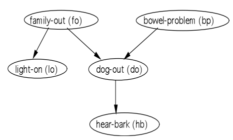
bp	¬bp

P(DO|FO,BP)

		do	¬do
fo	bp		
	¬bp		
¬fo	bp		
	¬bp		

P(FO|LO)

	lo	¬lo
fo		
¬fo		



How do BNs help?

Joint Probability Table

		hb				¬hb			
		do		¬do		do		¬do	
		bp	¬bp	bp	¬bp	bp	¬bp	bp	¬ bp
fo	lo								
	¬ lo								
¬fo	lo								
	¬ lo								

vs

Conditional Probability Tables

P(HB|DO)

	hb
do	
¬do	

P(FO)

fo

P(BP)

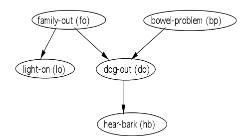
bp

P(DO|FO,BP)

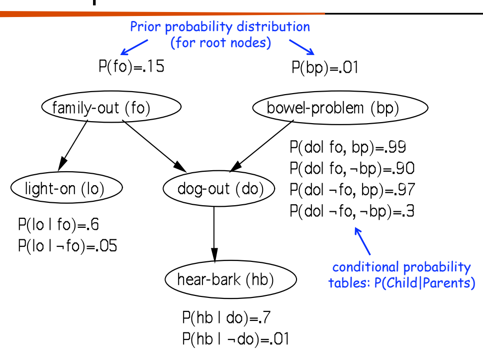
		do
fo	bp	
	¬bp	
¬fo	bp	
	¬bp	

P(FO|LO)

	lo
fo	
¬fo	



BN Example



Why don't these sum to 1?

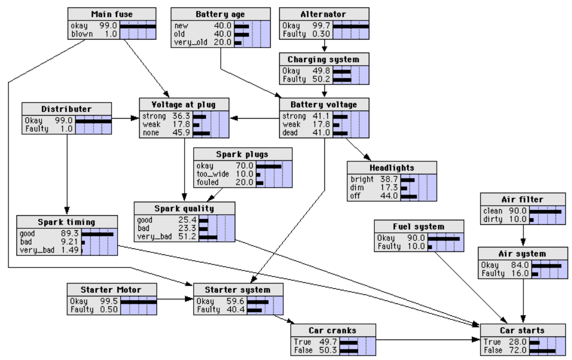
Bayes nets: Compactness

- How many numbers are required to build a Bayes Net
  - For a Boolean variable *X* with *k* Boolean parents, how many rows in the CPT?
    - 2^k
  - If each variable has no more than *k* parents and there are *n* nodes in the network, how many numbers required?
    - $n2^k$
- How many numbers required to specify the full joint distribution?
  - 2^n

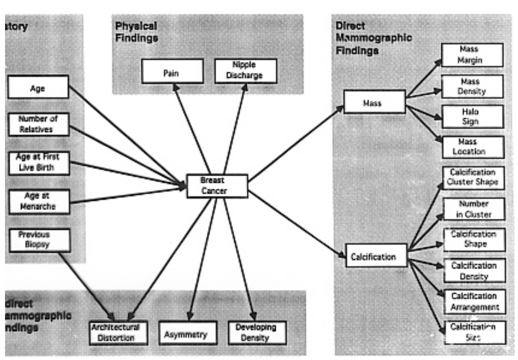
Bayes nets: Intuitiveness

- Can you estimate
  - P(do, lo, fo, bp, hb)?
  - P(do, lo, fo | bp, hb)?
  - …
- How about P(lo | fo) (lights out given that the family is out)?

Example: Car Diagnosis



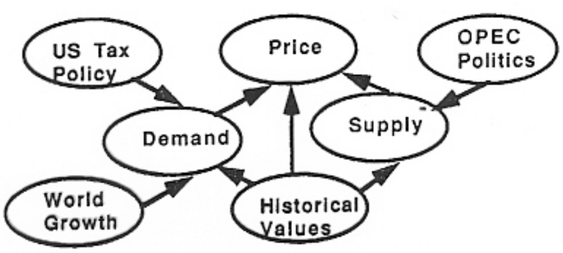
MammoNet: 88% accuracy



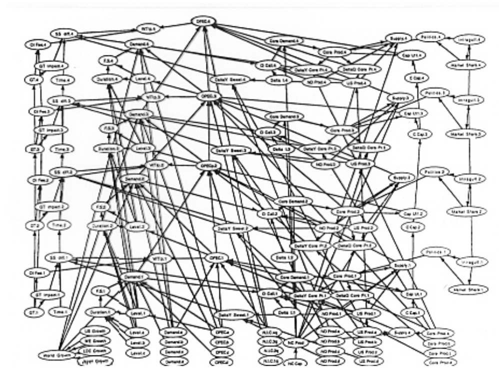
Other medical networks

- Mostly manually generated
  - PATHFINDER: pathology
  - MUNIN: neuromuscular disorders
  - CPCS (Computer-based Patient Case Study): internal medicine
    - 448 nodes
    - 8,254 conditional probability values
- Automatically generated
  - 100K to millions of nodes
    - e.g. text processing

ARCO1: Forecasting Oil Prices



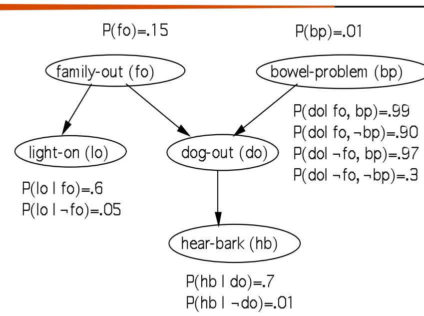
ARCO1: Forecasting Oil Prices



Asking questions about distributions

- We want to be able to ask questions about these probability distributions
- Given *n* variables, a query splits the variables into three sets:
  - query variable(s)
  - known/evidence variables
  - unknown/hidden variables
- P(query | evidence)
  - if we had no hidden variables, we could just multiply all the values in the different CPTs
  - to answer this, we need to sum over the hiden variables!

BN Example



p(fo | hb, lo)?

p(fo | hb, lo)

$$p(fo \mid hb, lo) = \frac{p(fo, hb, lo)}{p(hb, lo)}$$

Evidence: HB, LO
Query: FO
Hidden: BP, DO

$$p(FO \mid hb, lo) = \alpha\, p(FO, hb, lo)$$

$$= \alpha \sum_{bp} \sum_{do} p(FO, hb, lo, bp, do)$$

$$= \alpha \sum_{bp} \sum_{do} p(FO)\, p(bp)\, p(lo \mid FO)\, p(do \mid FO, bp)\, p(hb \mid do)$$

Any optimizations?

p(fo | hb, lo)

$$p(fo \mid hb, lo) = \frac{p(fo, hb, lo)}{p(hb, lo)}$$

Evidence: HB, LO
Query: FO
Hidden: BP, DO

$$p(FO \mid hb, lo) = \alpha\, p(FO, hb, lo)$$

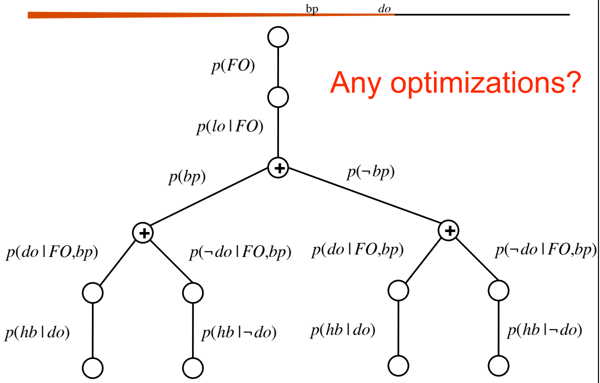
$$= \alpha \sum_{bp} \sum_{do} p(FO, hb, lo, bp, do)$$

$$= \alpha \sum_{bp} \sum_{do} p(FO)\, p(bp)\, p(lo \mid FO)\, p(do \mid FO, bp)\, p(hb \mid do)$$

$$= \alpha\, p(FO)\, p(lo \mid FO) \sum_{bp} p(bp) \sum_{do} p(do \mid FO, bp)\, p(hb \mid do)$$

$$p(FO \mid hb, lo) = \alpha\, p(FO)\, p(lo \mid FO) \sum_{bp} p(bp) \sum_{do} p(do \mid FO, bp)\, p(hb \mid do)$$

Any optimizations?



$$p(FO \mid hb, lo) = \alpha\, p(FO)\, p(lo \mid FO) \sum_{bp} p(bp) \sum_{do} p(do \mid FO, bp)\, p(hb \mid do)$$

Idea: calculate from the bottom up