CS 188: Artificial Intelligence
Spring 2010

Lecture 16: Bayes' Nets III – Inference
3/11/2010

Pieter Abbeel – UC Berkeley

Many slides over this course adapted from Dan Klein, Stuart Russell, Andrew Moore

---

$\forall i: P(x_i|x_1 \dots x_{i-1}) = P(x_i|Pa(X_i))$

Probabilities in BNs

- For all joint distributions, we have (chain rule):

$$P(x_1, x_2, \dots x_n) = \prod_{i=1}^{n} P(x_i|x_1, \dots, x_{i-1})$$

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \dots x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

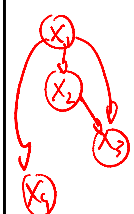
- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Announcements

- Current readings
  - Require login
- Assignments
  - W3 back today in lecture
  - W4 due tonight
- Midterm
  - 3/18, 6-9pm, 0010 Evans --- no lecture on 3/18
  - We will be posting practice midterms
  - One page note sheet, non-programmable calculators
  - Topics go through today, not next Tuesday

Example

$X_1, X_2, X_3, X_4, X_5, X_6$



$? \; P(X_2|X_1) \cancel{=} P(X_2)$

$P(X_3|X_1,X_2) \cancel{=} P(X_3|X_1)$

$\cancel{=} P(X_3|X_2)$

$P(X_4|X_1,X_2,X_3) = P(X_4|X_1) \stackrel{?}{=} P(X_4|X_2) \stackrel{?}{=} P(X_4|X_3)$

$\stackrel{?}{=} P(X_4|X_1,X_2) \stackrel{?}{=} P(X_4|X_2,X_3) \stackrel{?}{=} P(X_4|X_1,X_3)$

$P(X_4|X_5) \stackrel{?}{=} P(X_4)$

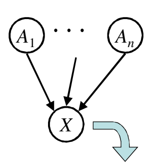
...

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \dots a_n)$$

  - CPT: conditional probability table
  - Description of a noisy "causal" process

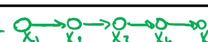


$P(X|A_1 \dots A_n)$

A Bayes net = Topology (graph) + Local Conditional Probabilities

---

$X_1 \perp X_5 | X_2$



$P(X_3|X_1,X_2) = P(X_3|X_2)$

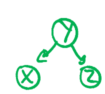
$P(X_4|X_1,X_2,X_3) = P(X_4|X_3)$

Conditional independence base cases

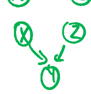
$P(X_5|X_1,\dots X_4) = P(X_5|X_4)$

- Causal chain — $X \perp Z | Y$

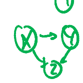
- Common cause — $X \perp Z | Y$



- Common effect — $X \perp Z \; ; \; \cancel{X \perp Z | Y}$ — $Y = XOR(X,Z)$



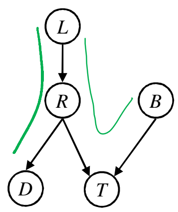
- Fully connected — $P(X) P(Y|X) P(Z|Y,X)$



- Fully disconnected

Reachability

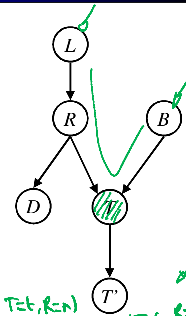
- Recipe: shade evidence nodes
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't count as a link in a path unless "active"



8

Example

$L \perp\!\!\!\perp T' \mid T$	Yes
$L \perp\!\!\!\perp B$	Yes
$L \perp\!\!\!\perp B \mid T$	No
$L \perp\!\!\!\perp B \mid T'$	
$L \perp\!\!\!\perp B \mid T, R$	Yes

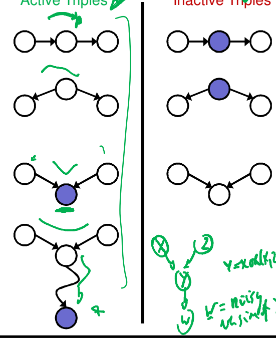


$P(L=l, B=b \mid T=t, R=r) = P(L=l \mid T=t, R=r) \cdot P(B=b \mid T=t, R=r)$

12

Reachability (D-Separation)

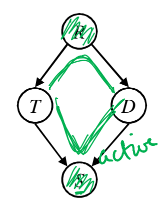
- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Yes, if X and Y "separated" by Z
  - Look for active paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain A → B → C where B is unobserved (either direction)
  - Common cause A ← B → C where B is unobserved
  - Common effect (aka v-structure) A → B ← C where B *or one of its descendents* is observed
- All it takes to block a path is a single inactive segment



Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad
- Questions:

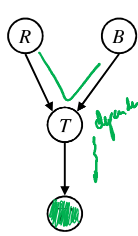
$T \perp\!\!\!\perp D$	No
$T \perp\!\!\!\perp D \mid R$	Yes
$T \perp\!\!\!\perp D \mid R, S$	No



13

Example

$R \perp\!\!\!\perp B$	Yes
$R \perp\!\!\!\perp B \mid T$	No
$R \perp\!\!\!\perp B \mid T'$	No



11

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - E.g. consider the variables *Traffic* and *Drips*
  - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology only guaranteed to encode conditional independence

$P(D \mid T) \neq P(D)$

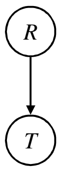
$T \rightarrow D$

$P(x_i \mid Pa(X_i)) = P(x_i \mid x_1, \ldots, x_{i-1})$

14

Example: Traffic

- Basic traffic net
- Let's multiply out the joint



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T,R)$

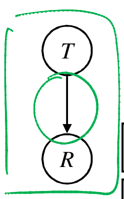
r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

$P(R|T) = \frac{P(R,T)}{P(T)}$

15

Example: Reverse Traffic

- Reverse causality?



$P(T)$

t	9/16
¬t	7/16

$P(R|T)$

t	r	1/3
	¬r	2/3
¬t	r	1/7
	¬r	6/7

$P(R|T) \neq P(R)$

$P(T,R)$

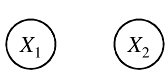
r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

16

Example: Coins

$P(X_4|X_1,X_2,X_3)$

- Extra arcs don't prevent representing independence, just allow non-independence

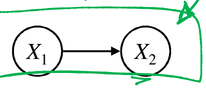


$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5



$P(X_1)$

h	0.5
t	0.5

$P(X_2|X_1)$

h | h	0.5
t | h	0.5
h | t	0.5
t | t	0.5

- Adding unneeded arcs isn't wrong, it's just inefficient

17

Changing Bayes' Net Structure

- The same joint distribution can be encoded in many different Bayes' nets
  - Causal structure tends to be the simplest
- Analysis question: given some edges, what other edges do you need to add?
  - One answer: fully connect the graph
  - Better answer: don't make any false conditional independence assumptions

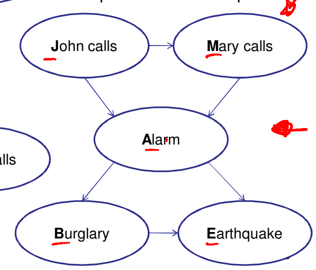
18

Example: Alternate Alarm

B ⫫ E

True distribution

If we reverse the edges, we make different conditional independence assumptions



To capture the same joint distribution, we have to add more edges to the graph

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution

20

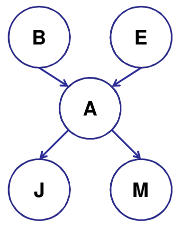
Inference

- Inference: calculating some useful quantity from a joint probability distribution
- Examples:
  - Posterior probability:

$$P(Q|E_1 = e_1, \ldots E_k = e_k)$$

  - Most likely explanation:

$$\text{argmax}_q \; P(Q = q|E_1 = e_1 \ldots)$$

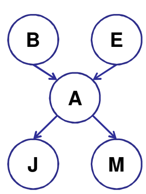


$P(+b,+j,+m,+a,+e) + P(+b,+j,+m,+a,\neg e) + P(+b,+j,+m,\neg a,\neg e)$
$+ P(+b,+j,+m,\neg a,+e)$

Inference by Enumeration

- Given unlimited time, inference in BNs is easy
- Recipe:
  - State the marginal probabilities you need
  - Figure out ALL the atomic probabilities you need
  - Calculate and combine them
- Example:

$$P(+b|+j,+m) = \frac{P(+b,+j,+m)}{P(+j,+m)}$$



Example: Enumeration

- In this simple method, we only need the BN to synthesize the joint entries

$$P(+b,+j,+m) =$$
$$P(+b)P(+e)P(+a|+b,+e)P(+j|+a)P(+m|+a)+$$
$$P(+b)P(+e)P(-a|+b,+e)P(+j|-a)P(+m|-a)+$$
$$P(+b)P(-e)P(+a|+b,-e)P(+j|+a)P(+m|+a)+$$
$$P(+b)P(-e)P(-a|+b,-e)P(+j|-a)P(+m|-a)$$

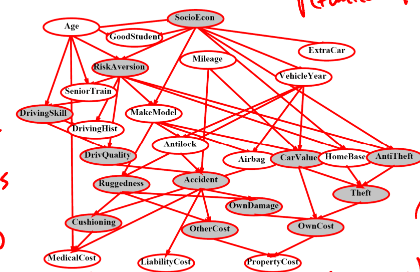
Inference by Enumeration?

$P(+\text{antilock} \mid e)$

2^{14}

terms

2^{14}



$\sum_u P(+\text{antilock}, u, e)$, $\sum_u P(\neg\text{antilock}, u, e)$