CS 188: Artificial Intelligence
Fall 2007

Lecture 16: Bayes Nets II

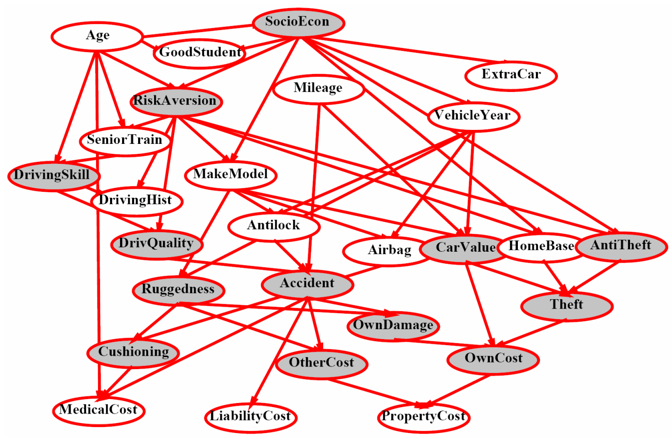
10/23/2007

Dan Klein – UC Berkeley

Bayes' Nets

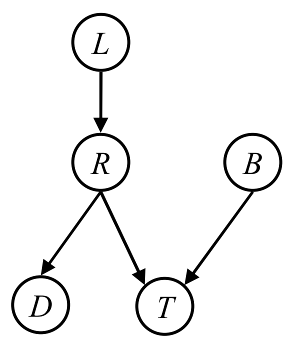
- A Bayes' net is an efficient encoding of a probabilistic model of a domain
- Questions we can ask:
  - Inference: given a fixed BN, what is P(X | e)?
  - Representation: given a fixed BN, what kinds of distributions can it encode?
  - Modeling: what BN is most appropriate for a given domain?

Example Bayes' Net



Example: Traffic

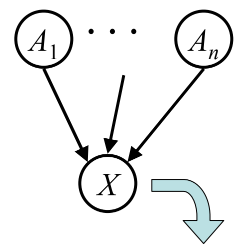
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame



Bayes' Net Semantics

- A Bayes' net:
  - A set of nodes, one per variable X
  - A directed, acyclic graph
  - A conditional distribution of each variable conditioned on its parents (the *parameters* θ)

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

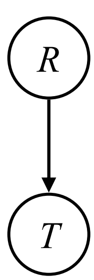


$$P(X|A_1 \ldots A_n)$$

- Semantics:
  - A BN defines a joint probability distribution over its variables:

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

Example: Mini Traffic



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

r →

t	3/4
¬t	1/4

¬r →

t	1/2
¬t	1/2

$$P(r, \neg t) =$$

Building the (Entire) Joint

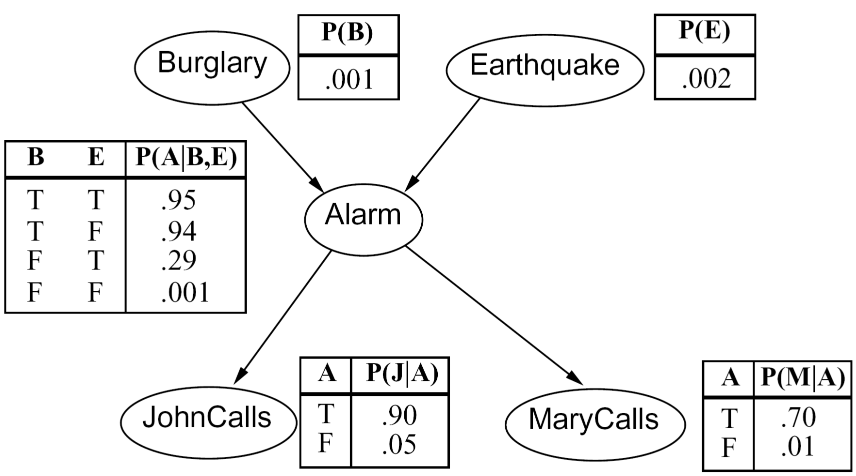
- We can take a Bayes' net and build any entry from the full joint distribution it encodes

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

  - Typically, there's no reason to build ALL of it
  - We build what we need on the fly

- To emphasize: every BN over a domain implicitly represents some joint distribution over that domain, but is specified by local probabilities

Example: Alarm Network



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

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

Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

- How big is an N-node net if nodes have k parents?

- Both give you the power to calculate $P(X_1, X_2, \dots X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (next class)

Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution

- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence

- After that: how to answer numerical queries (inference)

Conditional Independence

- Reminder: independence
  - X and Y are independent if

$$\forall x, y \;\; P(x,y) = P(x)P(y) \;\dashrightarrow\; X \perp\!\!\!\perp Y$$

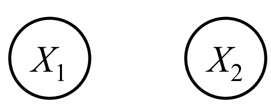
  - X and Y are conditionally independent given Z

$$\forall x, y, z \;\; P(x,y|z) = P(x)P(y) \;\dashrightarrow\; X \perp\!\!\!\perp Y | Z$$

  - (Conditional) independence is a property of a distribution

Example: Independence

- For this graph, you can fiddle with θ (the CPTs) all you want, but you won't be able to represent any distribution in which the flips are dependent!

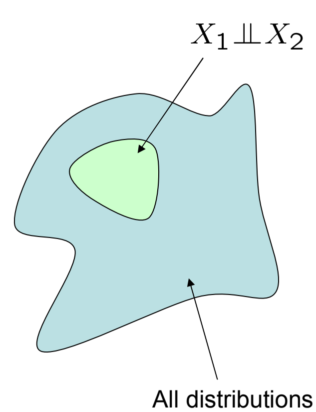


$P(X_1)$

h	0.5
t	0.5

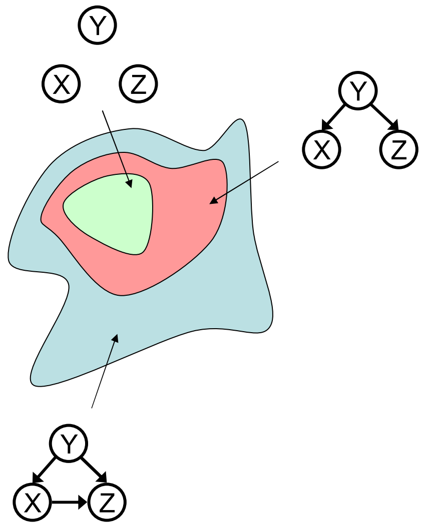
$P(X_2)$

h	0.5
t	0.5



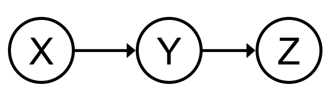
Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs



Independence in a BN

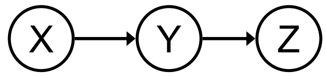
- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can calculate using algebra (really tedious)
  - If no, can prove with a counter example
  - Example:



  - Question: are X and Z independent?
    - Answer: not *necessarily*, we've seen examples otherwise: low pressure causes rain which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Addendum: they *could* be independent: how?

Causal Chains

- This configuration is a "causal chain"



X: Low pressure

Y: Rain

Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

  - Is X independent of Z given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

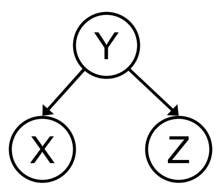
$$= P(z|y)$$

Yes!

  - Evidence along the chain "blocks" the influence

Common Cause

- Another basic configuration: two effects of the same cause
  - Are X and Z independent?
  - Are X and Z independent given Y?



$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

Yes!

Y: Project due

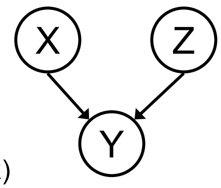
X: Newsgroup busy

Z: Lab full

  - Observing the cause blocks influence between effects.

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: remember the ballgame and the rain causing traffic, no correlation?
    - Still need to prove they must be (homework)
  - Are X and Z independent given Y?
    - No: remember that seeing traffic put the rain and the ballgame in competition?
  - This is backwards from the other cases
    - Observing the effect enables influence between effects.



X: Raining

Z: Ballgame

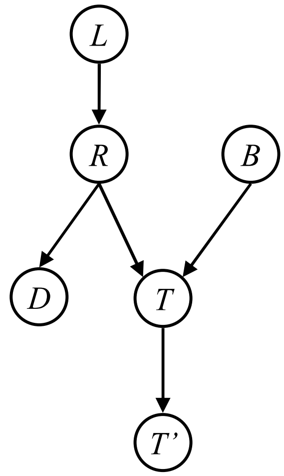
Y: Traffic

The General Case

- Any complex example can be analyzed using these three canonical cases
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: graph search!

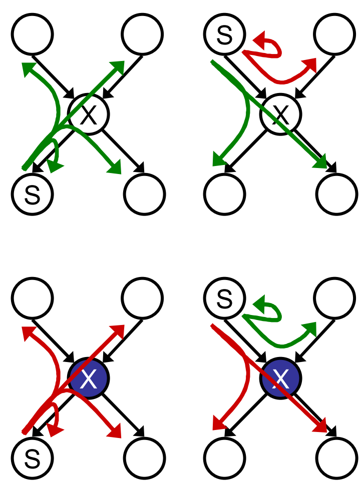
Reachability

- 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 shaded



Reachability (the Bayes' Ball)

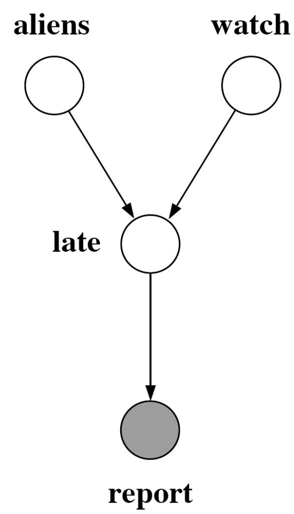
- Correct algorithm:
  - Shade in evidence
  - Start at source node
  - Try to reach target by search
  - States: pair of (node X, previous state S)
  - Successor function:
    - X unobserved:
      - To any child
      - To any parent if coming from a child
    - X observed:
      - From parent to parent
  - If you can't reach a node, it's conditionally independent of the start node given evidence



Example

$A \perp\!\!\!\perp W$ *Yes*

$A \perp\!\!\!\perp W | R$



Example

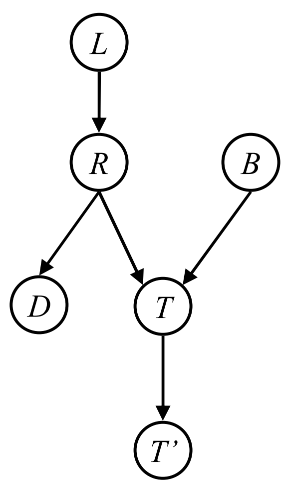
$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ *Yes*



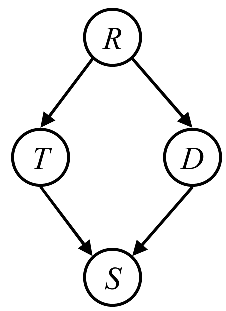
Example

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

  $T \perp\!\!\!\perp D$

  $T \perp\!\!\!\perp D | R$ Yes

  $T \perp\!\!\!\perp D | R, S$

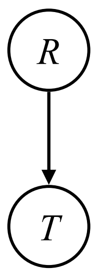


Causality?

- 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 independencies

Example: Traffic

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



$P(R)$

r	1/4
¬r	3/4

$P(T|R)$

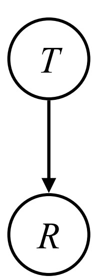
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T, R)$

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

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

$P(T, R)$

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

Example: Coins

- 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

$P(X_1)$

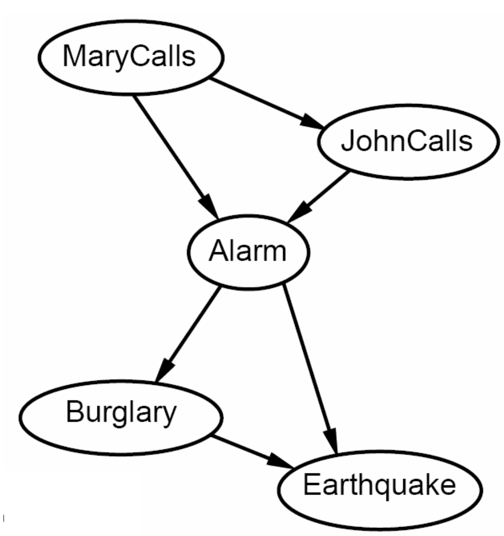
h	0.5
t	0.5

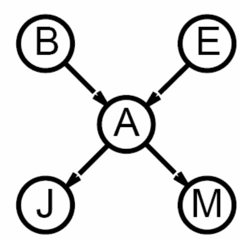
$P(X_2|X_1)$

h | h	0.5
t | h	0.5

h | t	0.5
t | t	0.5

Alternate BNs





Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- The Bayes' ball algorithm (aka d-separation)
- A Bayes' net may have other independencies that are not detectable until you inspect its specific distribution