

CS 188: Artificial Intelligence

Fall 2007

Lecture 15: Bayes Nets

10/18/2007

Dan Klein – UC Berkeley

Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

$$P(X_1, X_2, \dots, X_n)$$

- Given a joint distribution, we can reason about unobserved variables given observations (evidence)
- General form of a query:

$$\begin{array}{ccc} \text{Stuff you} & P(X_q | x_{e_1}, \dots, x_{e_k}) & \text{Stuff you} \\ \text{care about} & \xleftarrow{\hspace{1cm}} & \text{already know} \end{array}$$

- This kind of posterior distribution is also called the belief function of an agent which uses this model

Independence

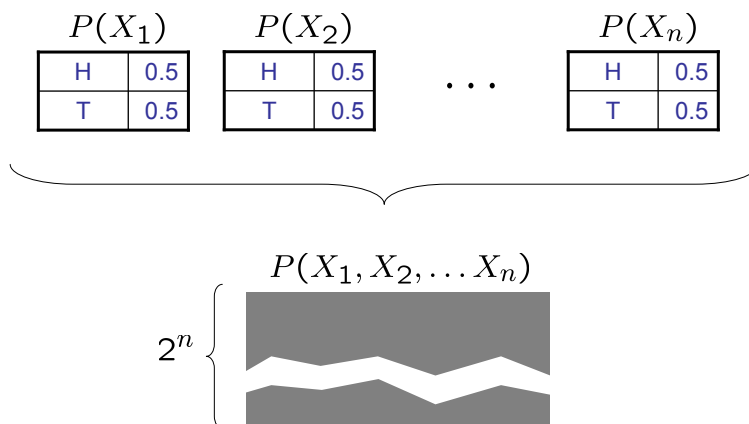
- Two variables are *independent* if:

$$P(X, Y) = P(X)P(Y)$$

- This says that their joint distribution *factors* into a product two simpler distributions
- Independence is a *modeling assumption*
 - Empirical* joint distributions: at best “close” to independent
 - What could we assume for {Weather, Traffic, Cavity, Toothache}?
- How many parameters in the joint model?
- How many parameters in the independent model?
- Independence is like something from CSPs: what?

Example: Independence

- N fair, independent coin flips:



Example: Independence?

- Most joint distributions are not independent
- Most are poorly modeled as independent

$$P(T)$$

T	P
warm	0.5
cold	0.5

$$P(S)$$

S	P
sun	0.6
rain	0.4

$$P(T, S)$$

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(T)P(S)$$

T	S	P
warm	sun	0.3
warm	rain	0.2
cold	sun	0.3
cold	rain	0.2

Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})?$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
 - $P(\text{catch} \mid \text{toothache}, \text{cavity}) = P(\text{catch} \mid \text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(\text{catch} \mid \text{toothache}, \neg \text{cavity}) = P(\text{catch} \mid \neg \text{cavity})$
- Catch is *conditionally independent* of Toothache given Cavity:
 - $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$
- Equivalent statements:
 - $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
 - $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$

Conditional Independence

- Unconditional (absolute) independence is very rare (why?)
- Conditional independence is our most basic and robust form of knowledge about uncertain environments:

$$P(X, Y|Z) = P(X|Z)P(Y|Z)$$

- What about this domain:
 - Traffic
 - Umbrella
 - Raining
- What about fire, smoke, alarm?

The Chain Rule II

- Can always factor any joint distribution as an incremental product of conditional distributions

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1) \dots$$

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i|X_1 \dots X_{i-1})$$

- Why?
- This actually claims nothing...
- What are the sizes of the tables we supply?

The Chain Rule III

- Trivial decomposition:

$$P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic})$$

- With conditional independence:

$$P(\text{Traffic, Rain, Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

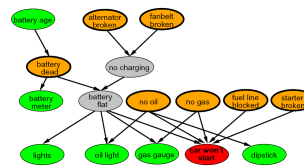
- Conditional independence is our most basic and robust form of knowledge about uncertain environments
- Graphical models help us manage independence

Graphical Models

- Models are descriptions of how (a portion of) the world works

- Models are always simplifications**

- May not account for every variable
- May not account for all interactions between variables



- What do we do with probabilistic models?

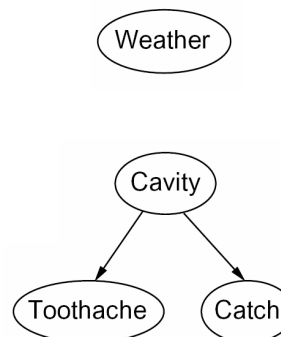
- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information

Bayes' Nets: Big Picture

- Two problems with using full joint distributions for probabilistic models:
 - Unless there are only a few variables, the joint is WAY too big to represent explicitly
 - Hard to estimate anything empirically about more than a few variables at a time
- Bayes' nets (more properly called graphical models) are a technique for describing complex joint distributions (models) using a bunch of simple, local distributions
 - We describe how variables locally interact
 - Local interactions chain together to give global, indirect interactions
 - For about 10 min, we'll be very vague about how these interactions are specified

Graphical Model Notation

- Nodes: variables (with domains)
 - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
 - Similar to CSP constraints
 - Indicate "direct influence" between variables
- For now: imagine that arrows mean causation



Example: Coin Flips

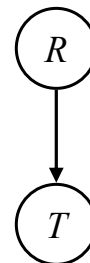
- N independent coin flips



- No interactions between variables:
absolute independence

Example: Traffic

- Variables:
 - R: It rains
 - T: There is traffic
- Model 1: independence
- Model 2: rain causes traffic
- Why is an agent using model 2 better?



Example: Traffic II

- Let's build a causal graphical model
- Variables
 - T: Traffic
 - R: It rains
 - L: Low pressure
 - D: Roof drips
 - B: Ballgame
 - C: Cavity

Example: Alarm Network

- Variables
 - B: Burglary
 - A: Alarm goes off
 - M: Mary calls
 - J: John calls
 - E: Earthquake!

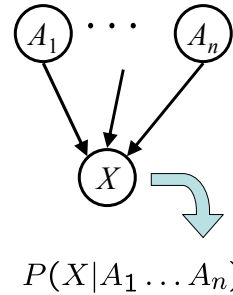
Bayes' Net Semantics

- 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

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



Probabilities in BNs

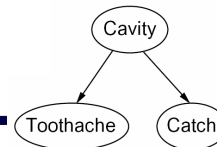
- Bayes' nets **implicitly** encode joint distributions
 - As a product of local conditional distributions
 - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

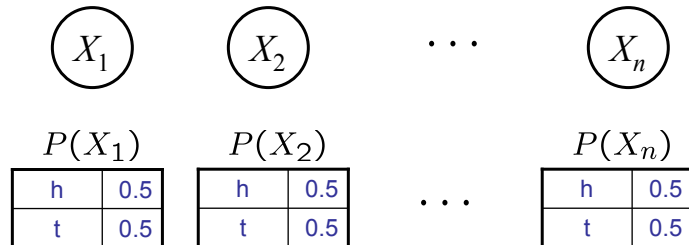
- Example:

$$P(\text{cavity}, \text{catch}, \neg \text{toothache})$$

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



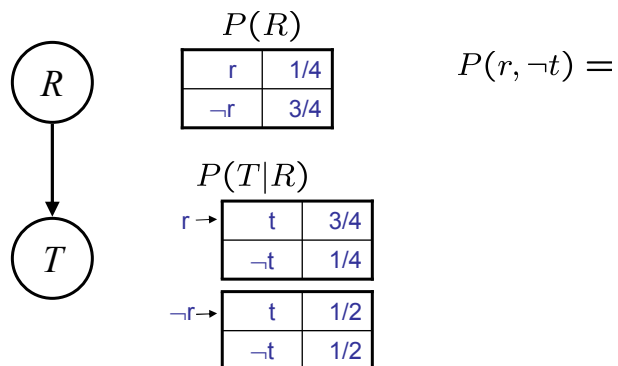
Example: Coin Flips



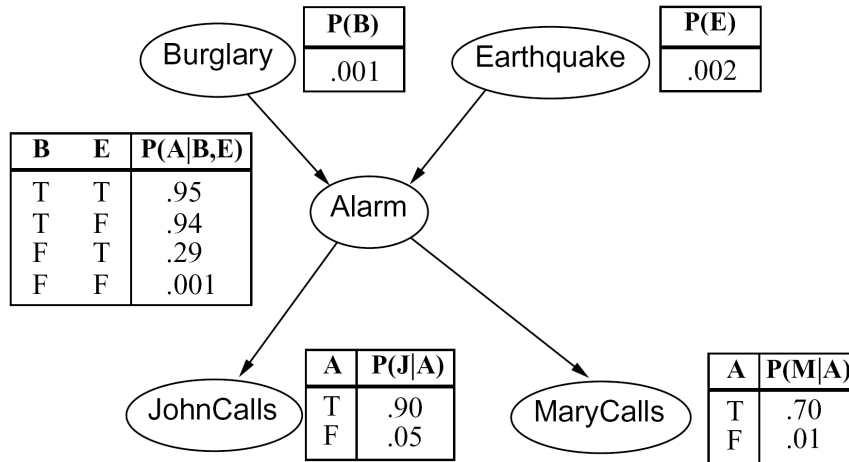
$P(h, h, t, h) =$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic



Example: Alarm Network



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

Example: Naïve Bayes

- Imagine we have one cause y and several effects x :

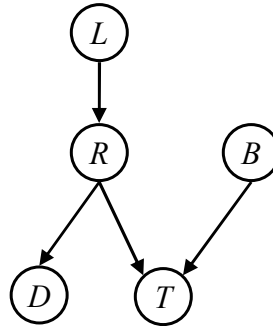
$$P(y, x_1, x_2 \dots x_n) = P(y)P(x_1|y)P(x_2|y) \dots P(x_n|y)$$

- This is a naïve Bayes model
- We'll use these for classification later

Example: Traffic II

- Variables

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



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)

Building the (Entire) Joint

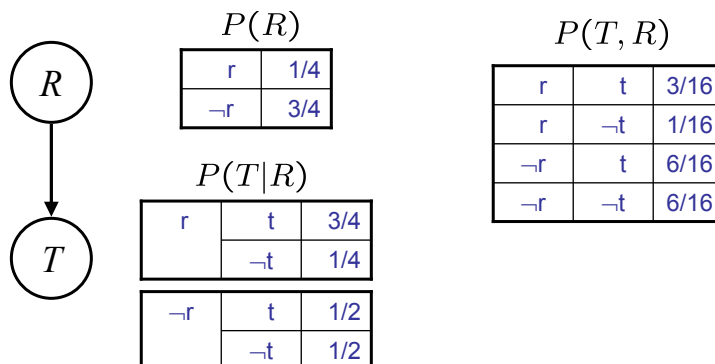
- We can take a Bayes' net and build the full joint distribution it encodes

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

- Typically, there's no reason to build ALL of it
 - But it's important to know you could!
- To emphasize: every BN over a domain **implicitly represents some joint distribution** over that domain

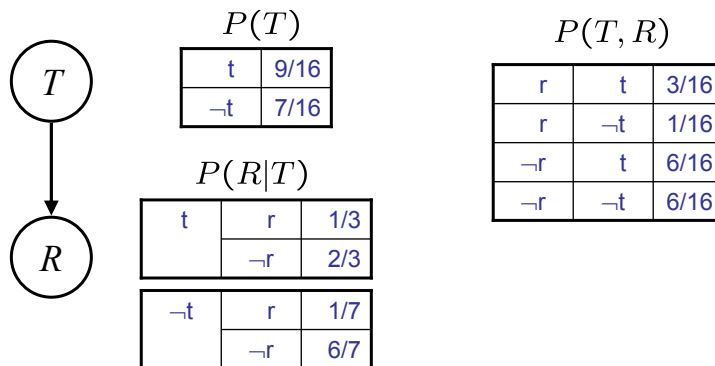
Example: Traffic

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



Example: Reverse Traffic

- Reverse causality?



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 (especially if variables are missing)
 - 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 really encodes conditional independencies

Creating Bayes' Nets

- So far, we talked about how any fixed Bayes' net encodes a joint distribution
- Next: how to represent a fixed distribution as a Bayes' net
 - Key ingredient: conditional independence
 - The exercise we did in "causal" assembly of BNs was a kind of intuitive use of conditional independence
 - Now we have to formalize the process
- After that: how to answer queries (inference)