Reasoning Under Uncertainty

Reading assignment:
"Bayesian Belief Networks"
by Scott Wooldridge

(Download from link on class website)

What is "Reasoning under Uncertainty"?

- Using probability theory to do automatic reasoning when not all aspects of the problem are known with certainty.
- E.g.:
  - Medical diagnosis
  - Speech recognition
  - Robot planning
  - Just about every AI task!

A famous example: The Monty Hall Problem

You are a contestant on a game show.

There are 3 doors, A, B, and C. There is a new car behind one of them and goats behind the other two.

Monty Hall, the host, asks you to pick a door, any door. You pick door A.

Monty tells you he will open a door , different from A, that has a goat behind it. He opens door B: behind it there is a goat.

Monty now gives you a choice: Stick with your original choice A or switch to C.

Should you switch?

http://math.ucsd.edu/~crypto/Monty/monty.html

Bayesian probability formulation

Hypothesis space H:

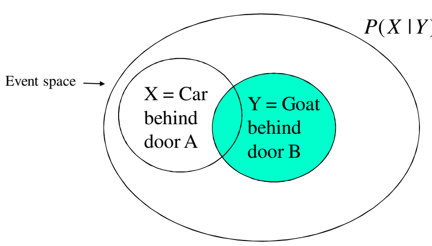
h_1 = Car is behind door A
h_2 = Car is behind door B
h_3 = Car is behind door C

Data D = Monty opened B

What is $P(h_1 | D)$?
What is $P(h_2 | D)$?
What is $P(h_3 | D)$?

Recall definition of conditional probability:

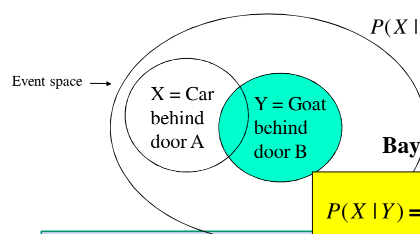
$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$



Event space = All possible configurations of cars and goats behind doors A, B, C

Recall definition of conditional probability:

$$P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$



Bayes Rule:

$$P(X | Y) = \frac{P(Y | X) P(X)}{P(Y)}$$

Proof :
$P(X | Y) P(Y) = P(X \cap Y) = P(Y | X) P(X)$

Terminology

- *Prior probability* of *h*:
  - *P(h)*: Probability that hypothesis *h* is true given our prior knowledge
  - If no prior knowledge, all $h \in H$ are equally probable
- *Posterior probability* of *h*:
  - *P(h|D)*: Probability that hypothesis *h* is true, given the data *D*.
- *Likelihood* of *D*:
  - *P(D|h)*: Probability that we will see data *D*, given hypothesis *h* is true.

Bayes rule says: $$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

Using Bayes' Rule to solve the Monty Hall problem

Data *D* = Monty opened door B

Hypothesis space *H*:
h_1 = Car is behind door A
h_2 = Car is behind door C
h_3 = Car is behind door B

What is P(h_1 | *D*)?
What is P(h_2 | *D*)?
What is P(h_2 | *D*)?

Prior probability:
P(h_1) = 1/3 P(h_2) =1/3 P(h_3) =1/3
Likelihood:
P(D | h_1) = 1/2
P(D | h_2) = 1
P(D | h_3) = 0

P(D) = p(D|h_1)p(h_1) + p(D|h_2)p(h_2) + p(D|h_3)p(h_3) = 1/6 + 1/3 + 0 = 1/2

By Bayes rule:
P(h1|D) = P(D|h1)p(h1) / P(D) = ½ 1/3 / ½ = 1/3

P(h2|D) = P(D|h2)p(h2) / P(D) = 1 1/3 / ½ = 2/3

So you should switch!

Another example
(Adapted from the reading)

A patient comes into a doctor's office with a bad cough and a high fever.

Hypothesis space *H*:
h_1: patient has flu
h_2: patient does not have flu

Data *D*:
coughing = true, *fever* = true

Prior probabilities:
p(h_1) = .1
p(h_2) = . 9

Likelihoods
p(***D*** | h1) = .8
p(***D*** | h2) = .4

Prob. of data
P(***D***) = P(D | h₁) P(h₁) + P(D | h₂) P(h₂) = .44

Posterior probabilities:
P(h_1|***D***) = P(D | h₁) P(h₁) / P(D) = .18

P(h_2|***D***) = .82

- Let's say we have the following random variables:

cough
fever
flu
smokes

Full joint probability distribution

smokes				
	cough		¬*cough*	
	Fever	¬*Fever*	*Fever*	¬*Fever*
flu	p_1	p_2	p_3	p_4
¬*flu*	p_5	p_6	p_7	p_8

¬*smokes*				
	cough		¬*cough*	
	fever	¬*fever*	*fever*	¬*fever*
flu	p_9	p_{10}	p_{11}	p_{12}
¬*flu*	p_{13}	p_{14}	p_{15}	p_{16}

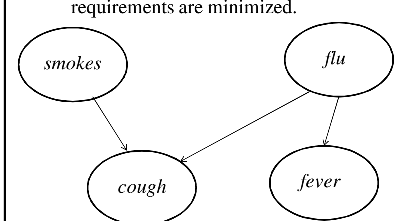
Sum of all boxes is 1.

In principle, the full joint distribution can be used to answer any question about probabilities of these combined parameters.

However, size of full joint distribution scales exponentially with number of parameters so is expensive to store and to compute with.

Bayesian networks

- Idea is to represent dependencies (or causal relations) for all the variables so that space and computation-time requirements are minimized.



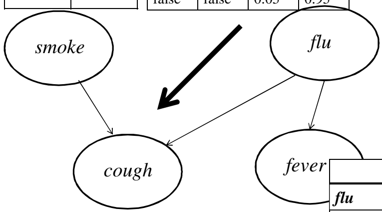
"Graphical Models"

Conditional probability tables for each node

smoke	
true	0.2
false	0.8

flu	smoke	cough true	cough false
True	True	0.95	0.05
True	False	0.8	0.2
False	True	0.6	0.4
false	false	0.05	0.95

flu	
true	0.01
false	0.99



flu	fever true	fever false
true	0.9	0.1
false	0.2	0.8

Semantics of Bayesian networks

- If network is correct, can calculate full joint probability distribution from network.

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$

where parents(X_i) denotes specific values of parents of X_i.

Example

- Calculate

$P(cough = t \wedge fever = f \wedge flu = f \wedge smoke = f)$

$$= \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$
$$= P(cough = t \mid flu = f \wedge smoke = f)$$
$$\times P(fever = f \mid flu = f)$$
$$\times P(flu = f)$$
$$\times P(smoke = f)$$
$$= .05 \times .8 \times .99 \times .8$$
$$= .032$$

Complexity of Bayesian Networks

For *n* random Boolean variables:

- Full joint probability distribution: 2^n entries
- Bayesian network with at most *k* parents per node:
  - Each conditional probability table: at most 2^k entries
  - Entire network: $n\,2^k$ entries

What are the advantages of Bayesian networks?

- Intuitive, concise representation of joint probability distribution (i.e., conditional dependencies) of a set of random variables.
- Represents “beliefs and knowledge” about a particular class of situations.
- Efficient (?) (approximate) inference algorithms
- Efficient, effective learning algorithms

Issues in Bayesian Networks

- Building / learning network topology
- Assigning / learning conditional probability tables
- Approximate inference via sampling