Artificial Intelligence

Uncertainty Reasoning

Non-monotonic Logic

- Traditional logic is monotonic
  - The set of legal conclusions grows monotonically with the set of facts appearing in our initial database
- When humans reason, we use defeasible logic
  - Almost every conclusion we draw is subject to reversal
  - If we find contradicting information later, we'll want to retract earlier inferences
- Nonmonotonic logic, or defeasible reasoning, allows a statement to be retracted
- Solution: Truth Maintenance
  - Keep explicit information about which facts/inferences support other inferences
  - If the foundation disappears, so must the conclusion

Uncertainty

- On the other hand, the problem might not be in the fact that T/F values can change over time but rather that we are not certain of the T/F value
- Agents almost never have access to the whole truth about their environment
- Agents must act in the presence of uncertainty
  - Some information ascertained from facts
  - Some information inferred from facts and knowledge about environment
  - Some information based on assumptions made from experience

Environment Properties

- Fully observable vs. partially observable
- Deterministic vs. stochastic / strategic
- Episodic vs. sequential
- Static vs. dynamic
- Discrete vs. continuous
- Single agent vs. multiagent

Uncertainty Arises Because of Several Factors

- Incompleteness
  - Many rules are incomplete because too many conditions to be explicitly enumerated
  - Many rules incomplete because some conditions are unknown
- Incorrectness

Where Do Probabilities Come From?

- Frequency
- Subjective judgment
- Consider the probability that the sun will still exist tomorrow.
- There are several ways to compute this
- Choice of experiment is known as the reference class problem

Acting Under Uncertainty

- Agents must still act even if world not certain
- If not sure which of two squares have a pit and must enter one of them to reach the gold, the agent will take a chance
- If can only act with certainty, most of the time will not act. Consider example that agent wants to drive someone to the airport to catch a flight, and is considering plan A90 that involves leaving home 60 minutes before the flight departs and driving at a reasonable speed. Even though the Pullman airport is only 5 miles away, the agent will not be able to reach a definite conclusion - it will be more like "Plan A90 will get us to the airport in time, as long as my car doesn't break down or run out of gas, and I don't get into an accident, and there are no accidents on the Moscow-Pullman highway, and the plane doesn't leave early, and there's no thunderstorms in the area, …"
- We may still use this plan if it will improve our situation, given known information
- The performance measure here includes getting to the airport in time, not wasting time at the airport, and/or not getting a speeding ticket.

Limitation of Deterministic Logic

- Pure logic fails for three main reasons:
- **Laziness**
  - Too much work to list complete set of antecedents or consequents needed to ensure an exceptionless rule, too hard to use the enormous rules that result
- **Theoretical ignorance**
  - Science has no complete theory for the domain
- **Practical ignorance**
  - Even if we know all the rules, we may be uncertain about a particular patient because all the necessary tests have not or cannot be run

Probability

- Probabilities are numeric values between 0 and 1 (inclusive) that represent ideal certainties (not beliefs) of statements, given assumptions about the circumstances in which the statements apply.
- These values can be verified by testing, unlike certainty values. They apply in highly controlled situations.

$$\text{Probability(event)} = \text{P(event)} = \frac{\text{\#instances of the event}}{\text{total \#instances}}$$

Example

- For example, if we roll two dice, each showing one of six possible numbers, the number of total unique rolls is 6*6 = 36. We distinguish the dice in some way (a first and second or left and right die). Here is a listing of the joint possibilities for the dice:

  (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
  (2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
  (3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
  (4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
  (5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
  (6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

- The number of rolls which add up to 4 is 3 ((1,3), (2,2), (3,1)), so the probability of rolling a total of 4 is 3/36 = 1/12.
- This does not mean 8.3% true, but 8.3% chance of it being true.

Probability Explanation

- P(event) is the probability in the absence of any additional information
- Probability depends on evidence.
- Before looking at dice: P(sum of 4) = 1/12
- After looking at dice: P(sum of 4) = 0 or 1, depending on what we see
- All probability statements must indicate the evidence with respect to which the probability is being assessed.
- As new evidence is collected, probability calculations are updated.
- Before specific evidence is obtained, we refer to the prior or unconditional probability of the event with respect to the evidence. After the evidence is obtained, we refer to the posterior or conditional probability.

Probability Distributions

- If we want to know the probability of a variable that can take on multiple values, we may define a probability distribution, or a set of probabilities for each possible variable value.
- TemperatureToday =

  {Below50, 50s, 60s, 70s, 80s, 90sAndAbove}
- P(TemperatureToday) =

  {0.1, 0.1, 0.5, 0.2, 0.05, 0.05}
- Note that the sum of the probabilities for possible values of any given variable must always sum to 1.

Joint Probability Distribution

- Because events are rarely isolated from other events, we may want to define a joint probability distribution, or $P(X_1, X_2, .., X_n)$.
- Each X_i is a vector of probabilities for values of variable X_i.
- The joint probability distribution is an n-dimensional array of combinations of probabilities.

	Wet	~Wet
Rain	**0.6**	**0.4**
~Rain	0.4	0.6

Inference by Enumeration

- To determine the probability of one variable (e.g., toothache), sum the events in the joint probability distribution where it is true:

	toothache		**~toothache**	
	catch	~catch	catch	~catch
cavity	.108	.012	.072	.008
~cavity	.016	.064	.144	.576

P (toothache) = .108 + .012 + .016 + .064 = 0.2 or 20%

This process is called “Marginalization

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$$

P(toothache$\vee$cavity =

.20 + ??

.072 + .008

.28

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Problems ??

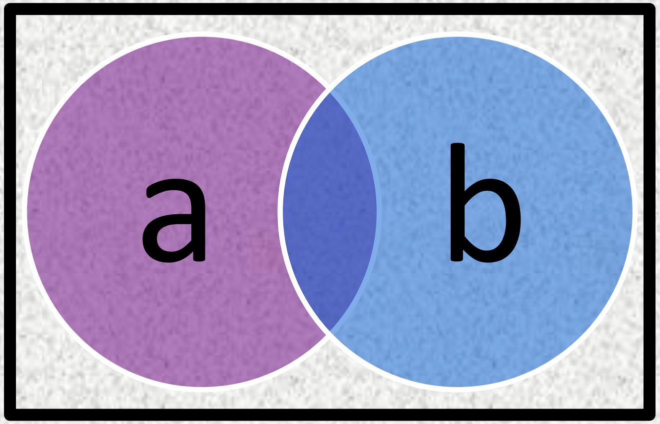
- Worst case time: $O(n^d)$
  - Where d = max arity
  - And n = number of random variables
- Space complexity also $O(n^d)$
  - Size of joint distribution
- How get $O(n^d)$ entries for table??

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Value of cavity & catch irrelevant - When computing P(toothache)

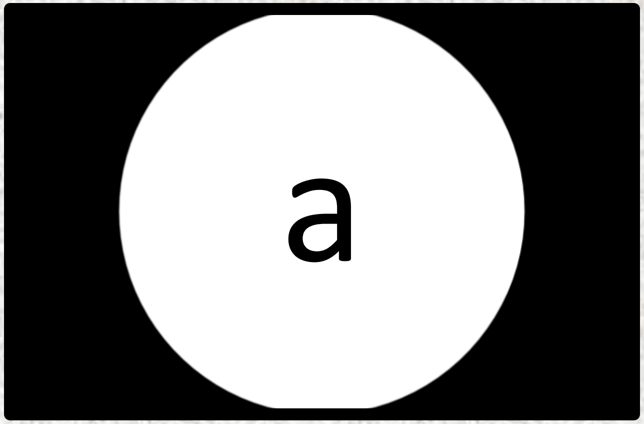
Axioms of Probability

- 0 <= P(Event) <= 1
- Disjunction, avb, P(avb) = P(a) + P(b) – P(a^b)



Axioms of Probability

- Negation, $P(\sim a) = 1 - P(a)$

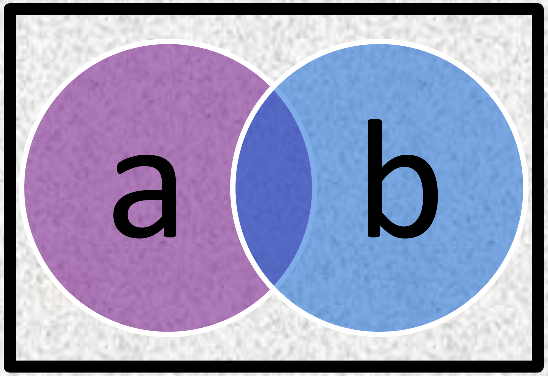


Axioms of Probability

- Conditional probability
  - Once evidence is obtained, the agent can use conditional probabilities, $P(a|b)$
  - $P(a|b)$ = probability of a being true given that we know b is true
  - The equation $P(a|b) = \dfrac{P(a \wedge b)}{P(b)}$ holds whenever $P(b)>0$
- An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome [deFinetti, 1931]

Axioms of Probability

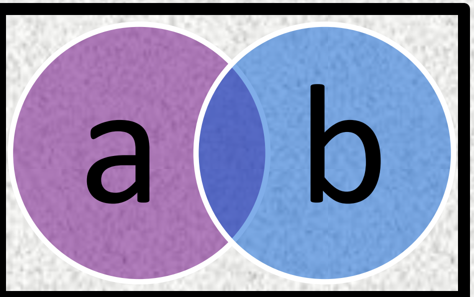
- Conjunction
  - Product rule
  - P(a^b) = P(a)*P(b|a)
  - P(a^b) = P(b)*P(a|b)

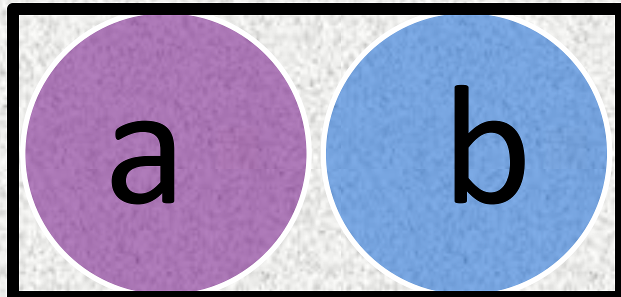


- In order words, the only way a and b can both be true is if a is true and we know b is true given a is true (thus b is also true)

Axioms of Probability

- If a and b are independent events (the truth of a has no effect on the truth of b), then $P(a \wedge b) = P(a) * P(b)$.
- “Wet” and “Raining” are not independent events.
- “Wet” and “Joe made a joke” are pretty close to independent events.





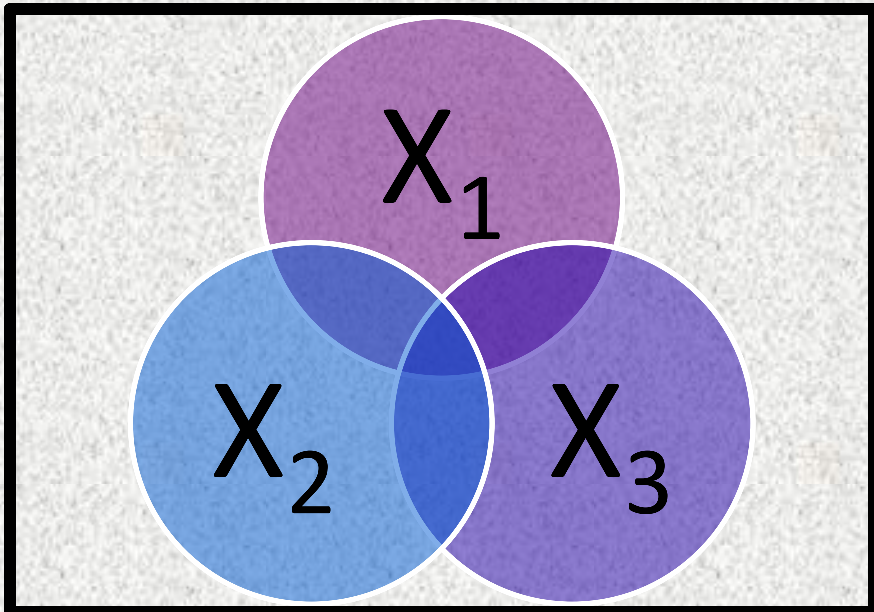
More Than 2 Variables

- The chain rule is derived by successive application of the product rule:
- $P(X_1,..,X_n) = P(X_1,..,X_{n-1})P(X_n|X_1,..,X_{n-1})$

  $= P((X_1,..,X_{n-2})P(X_{n-1}|X_1,..,X_{n-2})P(X_n|X_1,..,X_{n-1})$

  $= \ldots$

  $= \prod_{i=1}^{n} P(X_i|X_1,..,X_{i-1})$



Law of Alternatives

- If we know that exactly one of A1, A2, ..., An are true, then we know
  $P(B) = P(B|A1)P(A1) + P(B|A2)P(A2) + ... + P(B|An)P(An)$
  and
  $P(B|X) = P(B|A1,X) + ... + P(B|An,X)P(An,X)$
- Example
  - P(Sunday) = P(Monday) =.. = P(Saturday) = 1/7
  - P(FootballToday) =
    P(FootballToday|Sunday)P(Sunday) +
    P(FootballToday|Monday)P(Monday) +
    .. +
    P(FootballToday|Saturday)P(Saturday)
    = 0 + 0 + 0 + 0 + 0 + 0 + 1/7*1 = 1/7

Lunar Lander Example

- A lunar lander crashes somewhere in your town (one of the cells at random in the grid). The crash point is uniformly random (the probability is uniformly distributed, meaning each location has an equal probability of being the crash point).
- D is the event that it crashes downtown.
- R is the event that it crashes in the river.

					D	D	D	
R	R	R	R	R	DR	DR	DR	R
R	R	R	R	R	DR	DR	DR	R
					D	D	D	

What is P(R)? 18/54

What is P(D)? 12/54

What is P(D^R)? 6/54

What is P(D|R)? 6/18

What is P(R|D)? 6/12

What is P(R^D)/P(D)? 6/12

Axioms of Probability

- Bayes' Rule
  - Given a hypothesis (H) and evidence (E), and given that P(E) = 0, what is P(H|E)?
- Many times rules and information are uncertain, yet we still want to say something about the consequent; namely, the degree to which it can be believed.
- A British cleric and mathematician, Thomas Bayes, suggested an approach.
- Recall the two forms of the product rule:
  - P(ab) = P(a) * P(b|a)
  - P(ab) = P(b) * P(a|b)
- If we equate the two right-hand sides and divide by P(a), we get

$$P(b \mid a) = \frac{P(a \mid b) P(b)}{P(a)}$$

Example

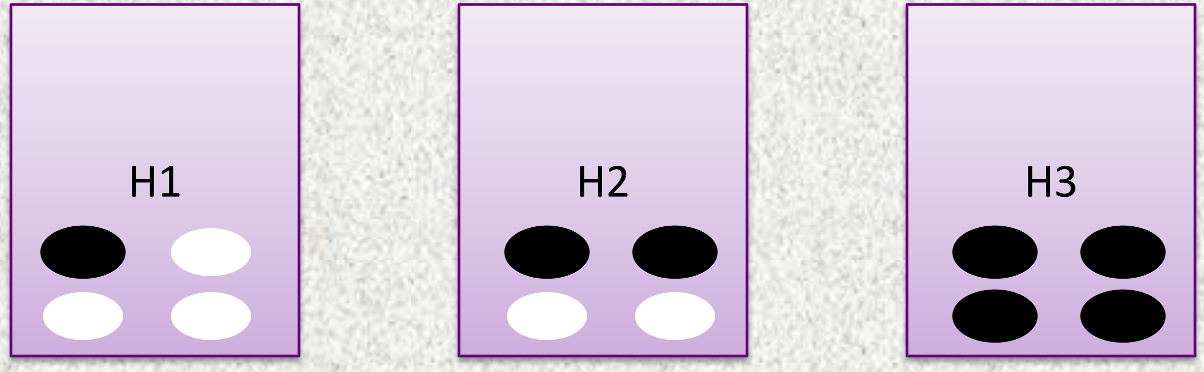
- Bayes' rule is useful when we have three of the four parts of the equation.
- In this example, a doctor knows that meningitis causes a stiff neck in 50% of such cases. The prior probability of having meningitis is 1/50,000 and the prior probability of any patient having a stiff neck is 1/20.
- What is the probability that a patient has meningitis if they have a stiff neck?
- H = "Patient has meningitis“
- E = "Patient has stiff neck"

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

$$P(H|E) = (0.5*.00002) / .05 = .0002$$

Submit : Oct 11, 2018

- I have three identical boxes labeled H1, H2, and H3
  - I place 1 black bead and 3 white beads into H1
  - I place 2 black beads and 2 white beads into H2
  - I place 4 black beads and no white beads into H3
- I draw a box at random, and randomly remove a bead from that box. Given the color of the bead, what can I deduce as to which box I drew?
- If I replace the bead, then redraw another bead at random from the same box, how well can I predict its color before drawing it?



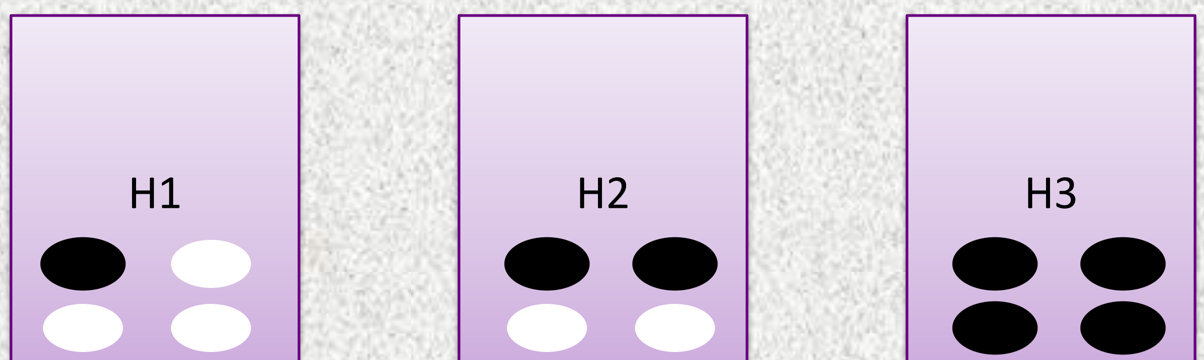
Answer

- Observation: I draw a white bead.
- $P(H1|W) = P(H1)P(W|H1) / P(W)$
  $= (1/3 * 3/4) / 5/12 = 3/12 * 12/5 = 36/60 = 3/5$
- $P(H2|W) = P(H2)P(W|H2) / P(W)$
  $= (1/3 * 1/2) / 5/12 = 1/6 * 12/5 = 12/30 = 2/5$
- $P(H3|W) = P(H3)P(W|H3) / P(W)$
  $= (1/3 * 0) / 5/12 = 0 * 12/5 = 0$

Example

- If I replace the bead, then redraw another bead at random from the same box, how well can I predict its color before drawing it?
- $P(H1)=3/5$, $P(H2) = 2/5$, $P(H3) = 0$
- $P(W) = P(W|H1)P(H1) + P(W|H2)P(H2) + P(W|H3)P(H3)$

$= 3/4*3/5 + 1/2*2/5 + 0*0 = 9/20 + 4/20 = 13/20$



Example

- We wish to know probability that John has malaria, given that he has a slightly unusual symptom: a high fever.
- We have 4 kinds of information
  a) probability that a person has malaria regardless of symptoms (0.0001)
  b) probability that a person has the symptom of fever given that he has malaria (0.75)
  c) probability that a person has symptom of fever, given that he does NOT have malaria (0.14)
  d) John has high fever
- H = John has malaria
- E = John has a high fever

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Suppose $P(H) = 0.0001$, $P(E|H) = 0.75$, $P(E|\sim H) = 0.14$

Example

- We wish to know probability that John has malaria, given that he has a slightly unusual symptom: a high fever.
- We have 4 kinds of information
  a) probability that a person has malaria regardless of symptoms (0.0001)
  b) probability that a person has the symptom of fever given that he has malaria (0.75)
  c) probability that a person has symptom of fever, given that he does NOT have malaria (0.14)
  d) John has high fever
- H = John has malaria
- E = John has a high fever

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Suppose P(H) = 0.0001, P(E|H) = 0.75, P(E|~H) = 0.14

Then P(E) = 0.75 * 0.0001 + 0.14 * 0.9999 = 0.14006

and P(H|E) = (0.75 * 0.0001) / 0.14006 = 0.0005354

On the other hand, if John did not have a fever, his probability of having malaria would be

$$P(H|\sim E) = \frac{P(\sim E|H) * P(H)}{P(\sim E)} = \frac{(1-0.75)(0.0001)}{(1-0.14006)} = 0.000029$$

Which is much smaller.

Making Decision Under Uncertainty

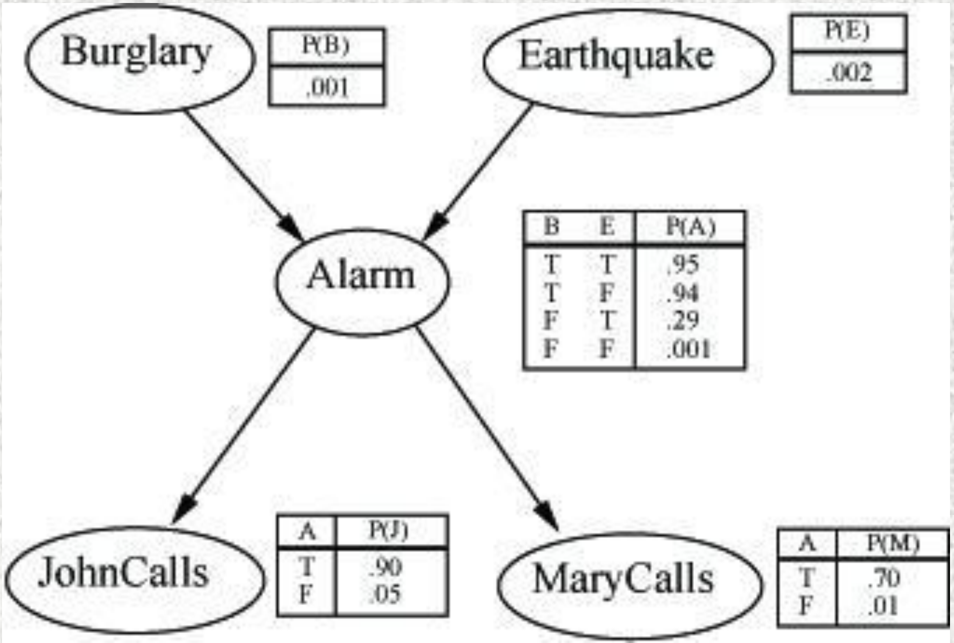
- Consider the following plans for getting to the airport:
  - $P(A_{25}$ gets me there on time | ...) = 0.04
  - $P(A_{90}$ gets me there on time | ...) = 0.70
  - $P(A_{120}$ gets me there on time | ...) = 0.95
  - $P(A_{1440}$ gets me there on time | ...) = 0.9999
- Which action should I choose?
- Depends on my preferences for missing the flight vs. time spent waiting, etc.
  - Utility theory is used to represent and infer preferences
  - Decision theory is a combination of probability theory and utility theory

Belief Networks

- A belief network (Bayes net) represents the dependence between variables.
- Components of a belief network graph:
- Nodes
  - These represent variables
- Links
  - X points to Y if X has a direct influence on Y
- Conditional probability tables
  - Each node has a CPT that quantifies the effects the parents have on the node
- The graph has no directed cycles

Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls
  Network topology reflects "causal" knowledge:

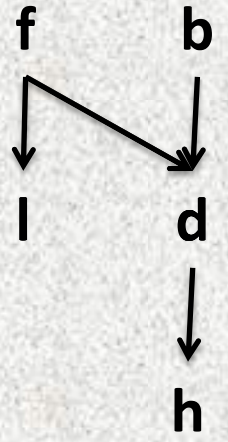


Example

- Suppose you are going home, and you want to know the probability that the lights are on given the dog is barking and the dog does not have a bowel problem. If the family is out, often the lights are on. The dog is usually in the yard when the family is out and when it has bowel troubles. If the dog is in the yard, it probably barks.
- Use the variables:
  f = family out
  l = light on
  b = bowel problem
  d = dog out
  h = hear bark
- There should be a graph with five nodes.

Example

- We know
  - l is directly influenced by f and is independent of b,d,h given f
    Add link from f to l
  - d is directly influenced by f and b, independent of l and h
    Add link from f to d and b to d
  - h is directly influenced by d, independent of f, l, b, and d
    Add link from d to h



Once we specify the topology (or learn it from data), we need to specify the conditional probability table for each node

p(f) = 0.15, 0.85
p(l|f) = 0.60, 0.40
p(d|f,b) = 0.99, 0.01
p(d|-f,b) = 0.97, 0.03
p(h|d) = 0.70, 0.30

p(b) = 0.01, 0.99
p(l|-f) = 0.05, 0.95
p(d|f,-b) = 0.90, 0.10
p(d|-f,-b) = 0.30, 0.70
p(h|-d) = 0.01, 0.99

Exercise

Consider the following Bayesian network, where F = having the flu and C = coughing:

P(F) = 0.1 $\quad F \longrightarrow C \quad$ P(C | F) = 0.8
P(C | ¬ F) = 0.3

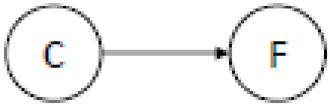
a) Write down the joint probability table specified by the Bayesian network.

Solution

Answer:

F	C	
t	t	$0.1 \times 0.8 = 0.08$
t	f	$0.1 \times 0.2 = 0.02$
f	t	$0.9 \times 0.3 = 0.27$
f	f	$0.9 \times 0.7 = 0.63$

b) Determine the probabilities for the following Bayesian network



so that it specifies the same joint probabilities as the given one.

Solution

Answer:

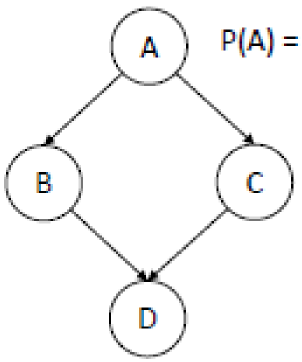
$P(C) = 0.08 + 0.27 = 0.35$

$P(F \mid C) = P(F, C)\ /\ P(C) = 0.08/0.35 \sim 0.23$

$P(F \mid \neg C) = P(F, \neg C)\ /\ P(\neg C) = 0.02/0.65 \sim 0.03$

Class exercise 2

Consider the following Bayesian network. A, B, C, and D are Boolean random variables. If we know that A is true, what is the probability of D being true?



$P(A) = 0.75$

$P(B \mid A) = 0.2$
$P(B \mid \neg A) = 0.5$

$P(C \mid A) = 0.7$
$P(C \mid \neg A) = 0.25$

$P(D \mid B \wedge C) = 0.3$
$P(D \mid B \wedge \neg C) = 0.25$
$P(D \mid \neg B \wedge C) = 0.1$
$P(D \mid \neg B \wedge \neg C) = 0.35$

Answer:

$\mathrm{P}(D|A)$ = P(A, D) / P(A)

= (P(A, B, C, D) + P(A, B, ¬ C, D) + P(A, ¬ B, C, D) + P(A, ¬ B, ¬ C, D)) / P(A)

= P(B | A) P(C | A) P(D | B, C) + P(B | A) P(¬ C | A) P(D | B, ¬ C) +

P(¬ B | A) P(C | A) P(D | ¬ B, C) + P(¬ B | A) P(¬ C | A) P(D | ¬ B, ¬ C)

$= (0.2 \times 0.7 \times 0.3) + (0.2 \times 0.3 \times 0.25) + (0.8 \times 0.7 \times 0.1) + (0.8 \times 0.3 \times 0.35)$

$= 0.042 + 0.015 + 0.056 + 0.084$

$= 0.197$