Probabilistic Reasoning

Bayesian Networks

- Directed Acyclic Graphs (DAGs)
- Arrows link Parents to Children
- Nodes represent random variables
- Each node knows P(this-node | parents(this-node))
  - Conditional Probability Tables
  - Sometimes programmed functions instead
- Arrow means direct influence, causes are parents of effects
  - Weather, Cavity, Toothache, Catch
  - Toothache and Catch are only conditionally independent given Cavity

Burglar alarms and earthquakes, Judea Pearl

- Burglary, P(B=true) = 0.001
- Earthquake, P(E=true) = 0.002
- Alarm, P(A=true | B, E) – note it isn't $\wedge$ this time

true	true	0.95
true	false	0.94
false	true	0.29
false	false	0.001

- JohnCalls, P(J=true | A)

true	0.90
false	0.05

- MaryCalls, P(M=true | A)

true	0.70
false	0.01

- For true/false variables no need to state P(X=false)
- No ListeningToLoudMusic node, some things we just don't know
- $P(J=true \wedge M=true \wedge A=true \wedge B=false \wedge E=false)$
  - $= P(J \mid A) \times P(M \mid A) \times P(A \mid \neg B \wedge \neg E) \times P(\neg B) \times P(\neg E)$
    - I left out all the "=true"s
  - $= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 \approx 0.000628$
  - We've effectively got the full joint distribution
- Direct influence only, Burglary does effect MaryCalls, but only indirectly
- Locally structured – never too many parents
  - Sometimes leave out very slight links
- Wrong order puts you in a bad position
  - Tables too big, probabilities difficult to discover
  - Go for Causal rather than Diagnostic

Utility Theory

- May be uncertain about which state the world is actually in
  - For each possible state S_i, we want a probability $P(S_i)$
- May be uncertain about actual outcomes of actions
  - $P(S_1 \rightarrow S_2, a)$
- Define $P(RESULT(a) = S_x) = \Sigma\, P(S_i) \times P(S_i \rightarrow S_x, a)$
- Utility function $U(S_x)$ is the utility (niceness) of being in state S_x
- Expected utility of an action $EU(a) = \Sigma\, P(RESULT(a) = S_x) \times U(S_x)$
- Maximum Expected Utility, MEU: action taken = argmax $EU(a_i)$
- Preferences: A, B two things that might be preferred one over the other

A > B - should be written curlily, the agent prefers A over B
A ~ B - Given a choice the agent wouldn't care
A >~ B - should be curly and on top, same as ¬(A > B)

- Lottery - set of possible outcomes with their probabilities
  [p_1,S_1; p_2,S_2; p_3,S_3; ...; p_N,S_N]
- Axioms of utility theory, for a rational agent (A and B are lotteries)
  Orderability: exactly one of A<B, A>B, A~B must be true
  Transitivity: (A>B) ∧ (B>C) ⇒ (A>C)
  Continuity: A > B > C ⇒ ∃p: [p,A; (1-p),C] ~ B
  Substitutability: A ~ B ⇒ [p,A; (1-p),C] ~ [p,B; (1-p),C]
  Monotonicity: A > B ⇒ (p>q ⇔ [p,A; (1-p),B] > [q,A; (1-q),B])
  Decomposability: [p,A; (1-p),[q,B; (1-q),C]] ~ [p,A; (1-p)×q,B; (1-p)×(1-q),C]
- If those axioms are obeyed, a sensible utility function U must exist
  U(A) > U(B) ⇔ A > B ∧ U(A) = U(B) ⇔ A ~ B
- Expected Utility of a lottery:
  U([p_1,S_1; p_2,S_2; p_3,S_3; ...; p_N,S_N]) = $\Sigma\, p_i \times U(S_i)$
- Utility Elicitation

Certainty Factors, from Mycin (pub. 1978)

- Mycin helped diagnose bacterial infections and recommend antibiotics
- Instead of Ture/False, every predicate has a certainty factor
- CF(P) = 1 means that P is certainly true
- CF(P) = 0 means we are completely ignorant about P
- CF(P) = -1 means that P is certainly false
- and intermediate values of course
- Basic facts come from observation and medical tests, all with CF's assigned
- CF($p_1 \wedge p_2 \wedge p_3 \wedge \ldots \wedge p_n$) = min(CF($p_i$))
- CF(¬p) = - CF(p)
- if CF(p) = x and CF(p ⇒ q) = y then CF(q) = x×y if x>0, or 0 otherwise
- Combining evidence: two rules lead to the same conclusion but with different CF's
  x and y are the two CF's:
  combined CF = x + y - x×y if x>=0 ∧ y>=0
  0 if x=0 ∧ y=0
  x + y + x×y if x<0 ∧ y<0
  (x+y) / (1 - min(abs(x), abs(y))) otherwise
- A tiny example:
  CF(joe-was-hungry) = 0.5
  CF(joe-was-hungry ⇒ joe-ate-something) = 0.8
  therefore CF(joe-ate-something) = 0.4
  CF(the-kitchen-is-messy ⇒ joe-ate-something) = 0.2
  but the kitchen is not messy,
  so CF(the-kitchen-is-messy) = -1.0,
  therefore CF(joe-ate-something) = 0.0
  All together CF(joe-ate-something) = (0.4 - 0.0) - (0.4 × 0.0) = 0.4
- The difference between that and probability?
  There is no probability that indicates ignorance,
  CF's apply to rules as well as facts.

Dempster-Shafer theory

- Another way of dealing with uncertainty
- Example: A person has been murdered and there are three suspects:

Miss Scarlett
Reverend Green
Colonel Mustard

- There are three possibilities:
  - MS - Miss Scarlett did it
  - RG - Reverend Green did it
  - CM - Colonel Mustard did it
  - exactly one of them must have done it
- We consider all eight possible subsets
  - A Mass (subjective probability) as assigned to each
    - All are equally likely: m({ MS }) = m({ RG }) = m({ CM }) = 0.3333
    - We have no idea means that m({ MS, RG, CM }) = 1.0
  - The subsets with non-zero mass are called Focal Elements
  - All the masses must add up to 1.0
  - Might have some direct evidence:
    - someone 60% sure she saw RG somewhere else, m({ MS, CM }) = 0.6
    - leaving m({ MS, RG, CM }) = 0.4, and all others = 0.0
  - Might have evidence that can't distinguish between individuals:
    - A witness is 80% sure a man did it, so m({ RG, CM }) = 0.8
    - leaving m({ MS, RG, CM }) = 0.2, and all others = 0.0
- Combining evidence, e.g. both of the examples above apply
  - $m_1(X)$ is the mass of X from case 1, $m_2(X)$ is from case 2
  - $m_C(X)$ is the combined mass from both pieces of evidence
    - $m_C(X)$ = sum of all possible $m_1(A) \times m_2(B)$, where $A \cap B = X$
  - $m_C(\{ CM \}) = 0.48$
  - $m_C(\{ MS, CM \}) = 0.12$
  - $m_C(\{ RG, CM \}) = 0.32$
  - $m_C(\{ MS, RG, CM \}) = 0.08$
  - all others are 0.0
- Belief, B(X) is the sum of the masses of all the subsets of X
  - $B_C(\{ CM \}) = 0.48$
  - $B_C(\{ MS, CM \}) = 0.6$, that is $m_C(\{ MS \}) + m_C(\{ CM \}) + m_C(\{ MS, CM \})$
  - $B_C(\{ RG, CM \}) = 0.8$, that is $m_C(\{ RG \}) + m_C(\{ CM \}) + m_C(\{ RG, CM \})$
  - $B_C(\{ MS, RG, CM \}) = 1.0$
  - all others are 0.0
- Plausibility, P(X) is $1 - B(\neg X)$
  - the sum of the masses of everything that contradicts X: it $\cap X = \emptyset$
  - $P_C(\{ CM \}) = 1.0$
  - $P_C(\{ MS, CM \}) = 1.0$
  - $P_C(\{ RG \}) = 0.4$
  - etc.
- Belief and Plausibility are claimed to be lower and upper bounds on probability
  - Belief(X) <= Probability(X) <= Plausibility(X)

Fuzzy Sets

- Represented as a Membership Function: m(item) $\in [0, 1]$
- For example, the fuzzy set BigNumbers might have
  - m(0) = 0, m(1) = 0, m(5) = 0.01, m(100) = 0.95, m(1000000) = 1
  - but of course m is a total function on all the real numbers
- The complement of a fuzzy set uses 1 - m
- $A \subseteq B \Leftrightarrow mA(x) <= mB(X)$ for all X in the domain
- The union of two sets is given by the maximum value of their m functions

- The intersection of two sets is given by the minimum value of their m functions