Chapter 11
Uncertainty & Reasoning under the uncertainty

Instructor
LE Thanh Sach, Ph.D.
Instructor’s Information

LE Thanh Sach, Ph.D.

Office:
Department of Computer Science,
Faculty of Computer Science and Engineering,
HoChiMinh City University of Technology.

Office Address:
268 LyThuongKiet Str., Dist. 10, HoChiMinh City, Vietnam.

E-mail: LTSACH@cse.hcmut.edu.vn
E-home: http://cse.hcmut.edu.vn/~ltsach/
Tel: (+84) 83-864-7256 (Ext: 5839)
Acknowledgment

The slides in this PPT file are composed using the materials supplied by

Prof. Stuart Russell and Peter Norvig: They are currently from University of California, Berkeley. They are also the author of the book “Artificial Intelligence: A Modern Approach”, which is used as the textbook for the course

Prof. Tom Lenaerts, from Université Libre de Bruxelles
Outline

- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule
- Bayesian Networks
Uncertainty

Let action $A_t = \text{leave for airport } t \text{ minutes} \text{ before flight}

Will $A_t$ get me there on time?

Problems:
1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Hence a purely logical approach either
1. risks falsehood: “$A_{25}$ will get me there on time”, or
2. leads to conclusions that are too weak for decision making:

“$A_{25}$ will get me there on time IF there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc.”

($A_{1440}$ might reasonably be said to get me there on time BUT I'd have to stay overnight in the airport …)
Methods for handling uncertainty

- Default or nonmonotonic logic:
  - Assume my car does not have a flat tire
  - Assume $A_{25}$ works unless contradicted by evidence
- Issues: What assumptions are reasonable? How to handle contradiction?

- Rules with fuzzy factors:
  - $A_{25} \rightarrow 0.3$ get there on time
  - $Sprinkler \rightarrow 0.99$ WetGrass
  - $WetGrass \rightarrow 0.7$ Rain
- Issues: Problems with combination, e.g., Sprinkler causes Rain??

- Probability
  - Model agent's degree of belief
  - Given the available evidence,
  - $A_{25}$ will get me there on time with probability 0.04
Probability

Probabilistic assertions **summarize** effects of
- **laziness**: failure to enumerate exceptions, qualifications, etc.
- **ignorance**: lack of relevant facts, initial conditions, etc.

**Subjective** probability:
- Probabilities relate propositions to agent's own state of knowledge
  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$

These are **not** assertions about the world

Probabilities of propositions change with new evidence:
  e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$
Making decisions under uncertainty

Suppose I believe the following:

\begin{align*}
P(A_{25} \text{ gets me there on time | ...}) & = 0.04 \\
P(A_{90} \text{ gets me there on time | ...}) & = 0.70 \\
P(A_{120} \text{ gets me there on time | ...}) & = 0.95 \\
P(A_{1440} \text{ gets me there on time | ...}) & = 0.9999 \\
\end{align*}

- Which action to choose?
  
  Depends on my preferences for missing flight vs. time spent waiting, etc.
  
  - Utility theory is used to represent and infer preferences
  - Decision theory = probability theory + utility theory
Syntax

- Basic element: \textit{random variable}
  - Similar to propositional logic: possible worlds defined by assignment of values to random variables.
  - \textbf{Boolean} random variables
    - e.g., \textit{Cavity} (do I have a cavity?)
  - \textbf{Discrete} random variables
    - e.g., \textit{Weather} is one of \{sunny, rainy, cloudy, snow\}

- Domain values:
  - must be \textbf{exhaustive and mutually exclusive}
Syntax

- **Elementary proposition:**
  - constructed by assignment of a value to a random variable:
  - e.g., \( \text{Weather} = \text{sunny} \), \( \text{Cavity} = \text{false} \) (abbreviated as \( \neg \text{cavity} \))

- **Complex propositions:**
  - formed from elementary propositions and standard logical connectives
  - e.g., \( \text{Weather} = \text{sunny} \lor \text{Cavity} = \text{false} \)
Syntax

- **Atomic event:**
  - A complete specification of the state of the world about which the agent is uncertain.

E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

\[
\begin{align*}
(Cavity = false) \land (Toothache = false) \\
Cavity = false \land Toothache = true \\
Cavity = true \land Toothache = false \\
Cavity = true \land Toothache = true
\end{align*}
\]

- Atomic events are **mutually exclusive and exhaustive**
Axioms of probability

- For any propositions $A$, $B$
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \lor B) = P(A) + P(B) - P(A \land B)$
Prior probability

- **Prior** or unconditional probabilities of propositions
  - e.g., \( P(Cavity = \text{true}) = 0.1 \) and \( P(Weather = \text{sunny}) = 0.72 \)
  - correspond to belief **prior** to arrival of any (new) evidence

- Probability distribution gives values for all possible assignments:
  - Weather’s domain: \(<\text{sunny, rainy, cloudy, snow}>\)
  - \( P(Weather) = <0.72, 0.1, 0.08, 0.1> \)
    (normalized, i.e., sums to 1)
Prior probability

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables.
  \( P(Weather, Cavity) = \) a 4 x 2 matrix of values:

<table>
<thead>
<tr>
<th>Weather</th>
<th>sunny</th>
<th>rainy</th>
<th>cloudy</th>
<th>snow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity = true</td>
<td>0.144</td>
<td>0.02</td>
<td>0.016</td>
<td>0.02</td>
</tr>
<tr>
<td>Cavity = false</td>
<td>0.576</td>
<td>0.08</td>
<td>0.064</td>
<td>0.08</td>
</tr>
</tbody>
</table>

- Every question about a domain can be answered by the joint distribution.
Conditional probability

- Conditional or posterior probabilities
  - e.g., $P(\text{cavity} \mid \text{toothache}) = 0.8$
  - i.e., given that toothache is all I know

- (Notation for conditional distributions:
  $P(\text{Cavity} \mid \text{Toothache}) = 2$-element vector of 2-element vectors)

- If we know more, e.g., cavity is also given, then we have
  $P(\text{cavity} \mid \text{toothache, cavity}) = 1$

- New evidence may be irrelevant, allowing simplification, e.g.,
  $P(\text{cavity} \mid \text{toothache, sunny}) = P(\text{cavity} \mid \text{toothache}) = 0.8$

- This kind of inference, sanctioned by domain knowledge, is crucial
Conditional probability

Definition of conditional probability:

\[ P(a \mid b) = \frac{p(a \land b)}{p(b)} \]

Product rule gives an alternative formulation:

\[ P(a \land b) = P(a, b) = P(a) \times P(b \mid a) = P(b) \times P(a \mid b) \]
Conditional probability

- A general version holds for whole distributions, e.g.,
  \[ P(\text{Weather, Cavity}) = P(\text{Weather} \mid \text{Cavity}) \cdot P(\text{Cavity}) \]

  (View as a set of 4 × 2 equations, not matrix mult.)

- **Chain rule** is derived by successive application of product rule:

  \[
  P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1}) \times P(X_n \mid X_1, \ldots, X_{n-1}) \\
  = P(X_1, \ldots, X_{n-1}) \times P(X_{n-1} \mid X_1, \ldots, X_{n-2}) \times P(X_n \mid X_1, \ldots, X_{n-1}) \\
  = \ldots \\
  = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})
  \]
Inference by enumeration

Start with the joint probability distribution:

<table>
<thead>
<tr>
<th>cavity</th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>.108</td>
<td>.072</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.012</td>
<td>.008</td>
</tr>
<tr>
<td>catch</td>
<td>.016</td>
<td>.144</td>
</tr>
<tr>
<td>¬ catch</td>
<td>.064</td>
<td>.576</td>
</tr>
</tbody>
</table>

For any proposition $\phi$, sum the atomic events where it is true: $P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>(\neg) toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.012</td>
</tr>
<tr>
<td>(\neg)catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>(\neg)cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- For any proposition \(\varphi\), sum the atomic events where it is true: \(P(\varphi) = \sum_{\omega: \omega \models \varphi} P(\omega)\)

\[
P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2
\]
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>̌toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.12</td>
</tr>
<tr>
<td>̌catch</td>
<td>0.072</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.064</td>
</tr>
<tr>
<td>̌cavity</td>
<td>0.144</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- Can also compute conditional probabilities:

\[
P(\neg \text{cavity} \mid \text{toothache}) = \frac{P(\neg \text{cavity}, \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\]
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.144</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>0.064</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- Can also compute conditional probabilities:

\[
P(cavity | toothache) = \frac{P(cavity, toothache)}{P(toothache)}
\]

\[
= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064}
\]

\[
= 0.6
\]
Inference by enumeration

- Start with the joint probability distribution:

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>0.108</td>
<td>0.072</td>
</tr>
<tr>
<td>¬ catch</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>cavity</td>
<td>0.016</td>
<td>0.144</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>0.064</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- Can also compute conditional probabilities:

\[ P(\text{cavity} | \text{toothache}) + P(\neg \text{cavity} | \text{toothache}) = 1 \]
Normalization

Denominator can be viewed as a normalization constant $\alpha$

$$P(Cavity | toothache) = \alpha, \ P(Cavity,toothache)$$
$$= \alpha, [P(Cavity,toothache,catch) + P(Cavity,toothache,\neg\ catch)]$$
$$= \alpha, [<0.108, 0.016> + <0.012, 0.064>]$$
$$= \alpha, <0.12, 0.08>$$
$$= <0.6, 0.4>$$

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables
Typically, we are interested in
the posterior joint distribution of the query variables \( Y \)
given specific values \( e \) for the evidence variables \( E \)

Let the hidden variables be \( H = X - Y - E \)

Then the required summation of joint entries is done by summing out the hidden variables:
\[
P(Y \mid E = e) = aP(Y, E = e) = a\sum_h P(Y, E = e, H = h)
\]

- The terms in the summation are joint entries because \( Y, E \) and \( H \) together exhaust the set of random variables

- Obvious problems:
  1. Worst-case time complexity \( O(d^n) \) where \( d \) is the largest arity
  2. Space complexity \( O(d^n) \) to store the joint distribution
  3. How to find the numbers for \( O(d^n) \) entries?
Independence

- A and B are independent iff
  \[ P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B) \quad \text{or} \quad P(A, B) = P(A)P(B) \]

- \[ P(\text{Toothache, Catch, Cavity, Weather}) = P(\text{Toothache, Catch, Cavity})P(\text{Weather}) \]

- 32 entries reduced to 12; for \( n \) independent biased coins, \( O(2^n) \rightarrow O(n) \)

- Absolute independence powerful but rare

- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?
Conditional independence

- $\mathbf{P}(\text{Toothache, Cavity, Catch})$ has $2^3 - 1 = 7$ independent entries.

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  
  $\text{(1) } \mathbf{P}(\text{catch} \mid \text{toothache, cavity}) = \mathbf{P}(\text{catch} \mid \text{cavity})$  

- The same independence holds if I haven't got a cavity:
  
  $\text{(2) } \mathbf{P}(\text{catch} \mid \text{toothache, } \neg\text{cavity}) = \mathbf{P}(\text{catch} \mid \neg\text{cavity})$  

- $\text{Catch}$ is conditionally independent of $\text{Toothache}$ given $\text{Cavity}$:
  
  $\mathbf{P}(\text{Catch} \mid \text{Toothache, Cavity}) = \mathbf{P}(\text{Catch} \mid \text{Cavity})$  

- Equivalent statements:
  
  $\mathbf{P}(\text{Toothache} \mid \text{Catch, Cavity}) = \mathbf{P}(\text{Toothache} \mid \text{Cavity})$  

  $\mathbf{P}(\text{Toothache, Catch} \mid \text{Cavity}) = \mathbf{P}(\text{Toothache} \mid \text{Cavity}) \mathbf{P}(\text{Catch} \mid \text{Cavity})$
Conditional independence contd.

- Write out full joint distribution using chain rule:
  \[
P(\text{Toothache}, \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Catch}, \text{Cavity}) \cdot P(\text{Catch}, \text{Cavity})
  = P(\text{Toothache} | \text{Catch}, \text{Cavity}) \cdot P(\text{Catch} | \text{Cavity}) \cdot P(\text{Cavity})
  = P(\text{Toothache} | \text{Cavity}) \cdot P(\text{Catch} | \text{Cavity}) \cdot P(\text{Cavity})
  \]

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in \(n\) to linear in \(n\).

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.
Bayes' Rule

- **Product rule:**
  \[
  P(a \land b) = P(a, b) = P(a) \times P(b \mid a) = P(b) \times P(a \mid b)
  \]

- **Bayes' rule:**
  \[
  P(a \mid b) = \frac{P(b \mid a) \times P(a)}{P(b)}
  \]
  or in distribution form
  \[
  P(a \mid b) = \alpha \times P(b \mid a) \times P(a)
  \]
Bayes' Rule

- **Usefulness:**
  - For assessing diagnostic probability from causal probability:

\[
P(cause \mid effect) = \frac{P(effect \mid cause) \times P(cause)}{P(effect)}
\]
Bayes' Rule

*Usefulness:

**Example:**

- Let $M$ be meningitis, (cause)
  - One patient in 10'000 people
- Let $S$ be stiff neck: (effect)
  - Ten patients in 100 people
- $P(S|M)$: 80% people effected by meningitis have stiff neck

\[
P(M | S) = \frac{P(S | M) \times P(M)}{P(S)}
\]

\[
= \frac{0.8 \times 0.0001}{0.1}
\]

\[
= 0.0008
\]

*Note: posterior probability of meningitis still very small!*
Bayes' Rule and conditional independence

\[
P(Cavity \mid toothache \land catch) = \alpha P(toothache \land catch \mid Cavity) P(Cavity)
\]

\[
= \alpha P(toothache \mid Cavity) P(catch \mid Cavity) P(Cavity)
\]

- This is an example of a naïve Bayes model:
  \[
P(Cause, Effect_1, \ldots, Effect_n) = P(Cause) \prod_i P(Effect_i \mid Cause)
\]

- Total number of parameters is linear in \(n\)
Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

\[ P \left( X_i \mid \text{Parents} \left( X_i \right) \right) \]
Bayesian networks

In the simplest case, a conditional distribution represented as a conditional probability table (CPT) giving the distribution over $X_i$ for each combination of parent values.
Example

- Topology of network encodes conditional independence assertions:

  - *Weather* is independent of the other variables
  - *Toothache* and *Catch* are conditionally independent given *Cavity*
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: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call
Example contd.
Compactness

- A CPT for Boolean $X_i$ with $k$ Boolean parents has $2^k$ rows for the combinations of parent values.

- Each row requires one number $p$ for $X_i = true$ (the number for $X_i = false$ is just $1-p$).

- If each variable has no more than $k$ parents, the complete network requires $O(n \cdot 2^k)$ numbers.

- I.e., grows linearly with $n$, vs. $O(2^n)$ for the full joint distribution.

- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5-1 = 31$).
Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

e.g., $P(j \land m \land a \land \neg b \land \neg e)$

$$= P(j \mid a) P(m \mid a) P(a \mid \neg b, \neg e) P(\neg b) P(\neg e)$$
Constructing Bayesian networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to $n$
   - add $X_i$ to the network
   - select parents from $X_1, \ldots, X_{i-1}$ such that
     $$P(X_i | \text{Parents}(X_i)) = P(X_i | X_1, \ldots, X_{i-1})$$

This choice of parents guarantees:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})$$  \hspace{1cm} \text{(chain rule)}$$

$$= \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i))$$  \hspace{1cm} \text{(by construction)}$$
Example

- Suppose we choose the ordering $M, J, A, B, E$

\[ P(J \mid M) = P(J) \]
Example

- Suppose we choose the ordering $M, J, A, B, E$

\[
P(J \mid M) = P(J) \quad \text{No}
\]

\[
P(A \mid J, M) = P(A \mid J) \quad P(A \mid J, M) = P(A)\
\]
Example

- Suppose we choose the ordering $M, J, A, B, E$

\[
P(J | M) = P(J) \ ? \ No
\]
\[
P(A | J, M) = P(A | J) \ ? \ P(A | J, M) = P(A) \ ? \ No
\]
\[
P(B | A, J, M) = P(B | A) \ ?
\]
\[
P(B | A, J, M) = P(B) \ ?
\]
Example

- Suppose we choose the ordering M, J, A, B, E

\[ P(J \mid M) = P(J)? \quad \text{No} \]
\[ P(A \mid J, M) = P(A \mid J)? \quad P(A \mid J, M) = P(A)? \quad \text{No} \]
\[ P(B \mid A, J, M) = P(B \mid A)? \quad \text{Yes} \]
\[ P(B \mid A, J, M) = P(B)? \quad \text{No} \]
\[ P(E \mid B, A, J, M) = P(E \mid A)? \]
\[ P(E \mid B, A, J, M) = P(E \mid A, B)? \]
Example

- Suppose we choose the ordering M, J, A, B, E

\[
P(J \mid M) = P(J) \text{? No}
\]

\[
P(A \mid J, M) = P(A \mid J) \text{? } P(A \mid J, M) = P(A) \text{? No}
\]

\[
P(B \mid A, J, M) = P(B \mid A) \text{? Yes}
\]

\[
P(B \mid A, J, M) = P(B) \text{? No}
\]

\[
P(E \mid B, A, J, M) = P(E \mid A) \text{? No}
\]

\[
P(E \mid B, A, J, M) = P(E \mid A, B) \text{? Yes}
\]
Example contd.

- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed
Bayesian Networks - Reasoning

- Example
  - Consider problem: “block-lifting”
  - B: the battery is charged.
  - L: the block is liftable.
  - M: the arm moves.
  - G: the gauge indicates that the battery is charged
Bayesian Networks - Reasoning

\[ P(B) = 0.95 \]

\[ P(L) = 0.7 \]

\[ P(G|B) = 0.95 \]
\[ P(G|\neg B) = 0.1 \]

\[ P(M|B,L) = 0.9 \]
\[ P(M|B,\neg L) = 0.05 \]
\[ P(M|\neg B,L) = 0.0 \]
\[ P(M|\neg B,\neg L) = 0.0 \]
Bayesian Networks - Reasoning

- Again, pls note:
  \[ p(G,M,B,L) = p(G|M,B,L)p(M|B,L)p(B|L)p(L) \]
  \[ = p(G|B)p(M|B,L)p(B)p(L) \]

- Specification:
  - Traditional: 16 rows
  - BayessianNetworks: 8 rows – see previous page.
Bayesian Networks - Reasoning

- Reasoning: top-down

Example:
- If the block is liftable, compute the probability of arm moving.

- I.e., Compute $p(M | L)$
Bayesian Networks - Reasoning

- **Reasoning:** top-down

- **Solution:**
  
  Insert parent nodes:
  
  \[
  p(M|L) = p(M,B|L) + p(M,\neg B|L)
  \]

  Use chain rule:
  
  \[
  p(M|L) = p(M|B,L)p(B|L) + p(M|\neg B,L)p(\neg B|L)
  \]

  Remove independent node:
  
  \[
  p(B|L) = p(B) : \quad B \text{ does not have PARENT}
  \]
  
  \[
  p(\neg B|L) = p(\neg B) = 1 - p(B)
  \]
Bayesian Networks - Reasoning

- Reasoning: top-down
- Solution:
  
  \[ p(M|L) = p(M|B,L)p(B) + p(M|\neg B,L)(1 - p(B)) \]
  
  \[ = 0.9 \times 0.95 + 0.0 \times (1 - 0.95) \]
  
  \[ = 0.855 \]
Bayesian Networks - Reasoning

- Reasoning: bottom-up
  
  Example:
  - If the arm cannot move
  - Compute the probability that the block is not liftable.
  - I.e., Compute: $p(\neg L | \neg M)$
Bayesian Networks - Reasoning

- Reasoning: bottom-up

Use Bayesian Rule:

\[ p(\neg L \mid \neg M) = \frac{p(\neg M \mid \neg L) p(\neg L)}{p(\neg M)} \]

Compute top-down reasoning

\[ p(\neg M \mid \neg L) = 0.9525 \quad \text{exercise} \]

\[ p(\neg L) = 1 - p(L) = 1 - 0.7 = 0.3 \]

\[ p(\neg L \mid \neg M) = \frac{0.9525 \times 0.3}{p(\neg M)} = \frac{0.28575}{p(\neg M)} \]
Bayesian Networks - Reasoning

Reasoning: bottom-up

Compute the negation component:

\[ p(L \mid \neg M) = \frac{0.0595 \times 0.7}{p(\neg M)} = \frac{0.03665}{p(\neg M)} \]

We have

\[ p(\neg L \mid \neg M) + p(L \mid \neg M) = 1 \]
\[ \Rightarrow p(\neg M) = 0.3224 \]

\[ \Rightarrow p(\neg L \mid \neg M) = 0.88632 \]
Bayesian Networks - Reasoning

- Reasoning: explanation
  Example
  - If we know \( \neg B \) (the battery is not charged)
  - Compute \( p(\neg L \mid \neg B, \neg M) \)
Bayesian Networks - Reasoning

Reasoning: explanation

\[
p(\neg L | \neg B, \neg M) = \frac{p(\neg M, \neg B | \neg L) p(\neg L)}{p(\neg B, \neg M)} \\
= \frac{p(\neg M | \neg B, \neg L) p(\neg B | \neg L) p(\neg L)}{p(\neg B, \neg M)} \\
= \frac{p(\neg M | \neg B, \neg L) p(\neg B) p(\neg L)}{p(\neg B, \neg M)}, \text{ because } B, L \text{ are independent} \\
= \frac{[1 - p(M | \neg B, \neg L)] \times [1 - p(B)] \times [1 - p(L)]}{p(\neg B, \neg M)} \\
= \frac{[1 - 0.0] \times [1 - 0.95] \times [1 - 0.7]}{p(\neg B, \neg M)} \\
= \frac{0.015}{p(\neg B, \neg M)}
\]
Bayesian Networks - Reasoning

- Reasoning: explanation

\[
p(L | \neg B, \neg M) = \frac{p(\neg M, \neg B | L) p(L)}{p(\neg B, \neg M)}
\]

\[
= \frac{p(\neg M | \neg B, L) p(\neg B | L) p(L)}{p(\neg B, \neg M)}
\]

\[
= \frac{p(\neg M | \neg B, L) p(\neg B) p(L)}{p(\neg B, \neg M)}, \text{ because B,L are independent}
\]

\[
= \frac{[1 - p(M | \neg B, L)] \times [1 - p(B)] \times p(L)}{p(\neg B, \neg M)}
\]

\[
= \frac{[1-0.0] \times [1-0.95] \times 0.7}{p(\neg B, \neg M)}
\]

\[
= \frac{0.035}{p(\neg B, \neg M)}
\]
Bayesian Networks - Reasoning

- Reasoning: explanation

\[
p(\neg L \mid \neg B, \neg M) + p(L \mid \neg B, \neg M) = 1
\]

\[
\Rightarrow \frac{0.015}{p(\neg B, \neg M)} + \frac{0.035}{p(\neg B, \neg M)} = 1
\]

\[
\Rightarrow p(\neg B, \neg M) = 0.045
\]

\[
\Rightarrow p(\neg L \mid \neg B, \neg M) = \frac{0.015}{0.045} = 0.33
\]
Summary

- Probability is a rigorous formalism for uncertain knowledge
- **Joint probability distribution** specifies probability of every atomic event
- Queries can be answered by summing over atomic events
- For nontrivial domains, we must find a way to reduce the joint size
- **Independence** and **conditional independence** provide the tools
Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct