Uncertainty
Chapter 13

Mausam

(Based on slides by UW-AI faculty)
Knowledge Representation

<table>
<thead>
<tr>
<th>KR Language</th>
<th>Ontological Commitment</th>
<th>Epistemological Commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propositional Logic</td>
<td>facts</td>
<td>true, false, unknown</td>
</tr>
<tr>
<td>First Order Logic</td>
<td>facts, objects, relations</td>
<td>true, false, unknown</td>
</tr>
<tr>
<td>Temporal Logic</td>
<td>facts, objects, relations, times</td>
<td>true, false, unknown</td>
</tr>
<tr>
<td>Probability Theory</td>
<td>facts</td>
<td>degree of belief</td>
</tr>
<tr>
<td>Fuzzy Logic</td>
<td>facts, degree of truth</td>
<td>known interval values</td>
</tr>
</tbody>
</table>

Probabilistic Relational Models
- combine probability and first order logic
Need for Reasoning w/ Uncertainty

• The world is full of uncertainty
  – chance nodes/sensor noise/actuator error/partial info..
  – Logic is brittle
    • can’t encode exceptions to rules
    • can’t encode statistical properties in a domain
  – Computers need to be able to handle uncertainty

• Probability: new foundation for AI (& CS!)

• Massive amounts of data around today
  – Statistics and CS are both about data
  – Statistics lets us summarize and understand it
  – Statistics is the basis for most learning

• Statistics lets data do our work for us
## Logic vs. Probability

<table>
<thead>
<tr>
<th>Symbol: Q, R …</th>
<th>Random variable: Q …</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean values: T, F</td>
<td>Domain: you specify e.g. {heads, tails} [1, 6]</td>
</tr>
<tr>
<td>State of the world: Assignment to Q, R … Z</td>
<td>Atomic event: complete specification of world: Q… Z</td>
</tr>
<tr>
<td></td>
<td>• Mutually exclusive</td>
</tr>
<tr>
<td></td>
<td>• Exhaustive</td>
</tr>
<tr>
<td></td>
<td>Prior probability (aka Unconditional prob: P(Q))</td>
</tr>
<tr>
<td></td>
<td>Joint distribution: Prob. of every atomic event</td>
</tr>
</tbody>
</table>
Probability Basics

• Begin with a set $S$: the sample space
  – e.g., 6 possible rolls of a die.

• $x \in S$ is a sample point/possible world/atomic event

• A probability space or probability model is a sample space with an assignment $P(x)$ for every $x$ s.t. $0 \leq P(x) \leq 1$ and $\sum P(x) = 1$

• An event $A$ is any subset of $S$
  – e.g. $A$ = ‘die roll < 4’

• A random variable is a function from sample points to some range, e.g., the reals or Booleans
Types of Probability Spaces

Propositional or Boolean random variables
   e.g., Cavity (do I have a cavity?)

Discrete random variables (*finite* or *infinite*)
   e.g., Weather is one of (*sunny, rain, cloudy, snow*)
   Weather = rain is a proposition
   Values must be exhaustive and mutually exclusive

Continuous random variables (*bounded* or *unbounded*)
   e.g., Temp = 21.6; also allow, e.g., Temp < 22.0.

Arbitrary Boolean combinations of basic propositions
Axioms of Probability Theory

• All probabilities between 0 and 1
  – $0 \leq P(A) \leq 1$
  – $P(\text{true}) = 1$
  – $P(\text{false}) = 0$.

• The probability of disjunction is:

$$P(A \lor B) = P(A) + P(B) - P(A \land B)$$
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:
  \( P(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle \) (normalized, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the
  probability of every atomic event on those r.v.s
  \( P(Weather, Cavity) = \) a \( 4 \times 2 \) matrix of values:

Joint distribution can answer any question
Conditional probability

• **Conditional or posterior probabilities**
  
e.g., $P(\text{cavity} \mid \text{toothache}) = 0.8$
  
i.e., given that \text{toothache} is all I know there is 80% chance of cavity

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

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

• New evidence may be irrelevant, allowing simplification:
  
$P(\text{cavity} \mid \text{toothache, sunny}) = P(\text{cavity} \mid \text{toothache}) = 0.8$

• This kind of inference, sanctioned by domain knowledge, is crucial
Conditional Probability

- $P(A \mid B)$ is the probability of $A$ given $B$
- Assumes that $B$ is the only info known.
- Defined by:

$$P(A \mid B) = \frac{P(A \land B)}{P(B)}$$
Chain Rule/Product Rule

\[ P(X_1, ..., X_n) = P(X_n | X_1..X_{n-1})P(X_{n-1} | X_1..X_{n-2})... P(X_1) = \prod P(X_i | X_1,..X_{i-1}) \]
What is the probability of a cavity given a toothache?
What is the probability of a cavity given the probe catches?
Inference by Enumeration

Start with the joint distribution:

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

For any proposition $\phi$, sum the atomic events where it is true:

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

$$P(\text{toothache}) = .108 + .012 + .016 + .064 = .20 \text{ or } 20\%$$
Inference by Enumeration

Start with the joint distribution:

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

For any proposition $\phi$, sum the atomic events where it is true:

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

$$P(\text{toothache} \lor \text{cavity}) = .20 + .072 + .008 = .28$$
Inference by Enumeration

Start with the joint 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:

\[
P(\neg \text{cavity}|\text{toothache}) = \frac{P(\neg \text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\]
Complexity of Enumeration

• Worst case time: $O(d^n)$
  – Where $d = \text{max arity}$
  – And $n = \text{number of random variables}$

• Space complexity also $O(d^n)$
  – Size of joint distribution

• Prohibitive!
Independence

• A and B are *independent* iff:

\[
\begin{align*}
P(A | B) &= P(A) \\
P(B | A) &= P(B)
\end{align*}
\]

• Therefore, if A and B are independent:

\[
\begin{align*}
P(A | B) &= \frac{P(A \land B)}{P(B)} = P(A) \\
P(A \land B) &= P(A)P(B)
\end{align*}
\]
Independence

$A$ and $B$ are independent iff
$P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)P(B)$

Complete independence is powerful but rare
What to do if it doesn’t hold?
Conditional Independence

\[ P(\text{Toothache}, \text{Cavity}, \text{Catch}) \] 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:

(1) \[ P(\text{catch} | \text{toothache}, \text{cavity}) = P(\text{catch} | \text{cavity}) \]

The same independence holds if I haven’t got a cavity:

(2) \[ P(\text{catch} | \text{toothache}, \neg \text{cavity}) = P(\text{catch} | \neg \text{cavity}) \]

*Catch is conditionally independent of Toothache given Cavity:* \[ P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity}) \]

Instead of 7 entries, only need 5
Conditional Independence II

\[ P(\text{catch} \mid \text{toothache, cavity}) = P(\text{catch} \mid \text{cavity}) \]
\[ P(\text{catch} \mid \text{toothache, } \neg\text{cavity}) = P(\text{catch} \mid \neg\text{cavity}) \]

Equivalent statements:
\[ P(\text{Toothache} \mid \text{Catch, Cavity}) = P(\text{Toothache} \mid \text{Cavity}) \]
\[ P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})P(\text{Catch} \mid \text{Cavity}) \]

Why only 5 entries in table?

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

I.e., \(2 + 2 + 1 = 5\) independent numbers (equations 1 and 2 remove 2)
Power of Cond. Independence

• Often, using conditional independence reduces the storage complexity of the joint distribution from exponential to linear!!

• Conditional independence is the most basic & robust form of knowledge about uncertain environments.
Bayes Rule

$P(x, y) = P(x \mid y)P(y) = P(y \mid x)P(x)$

$P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$

Useful for assessing diagnostic probability from causal probability:

$P(\text{Cause} \mid \text{Effect}) = \frac{P(\text{Effect} \mid \text{Cause})P(\text{Cause})}{P(\text{Effect})}$
E.g. let $M$ be meningitis, $S$ be stiff neck

$P(M) = 0.0001,$
$P(S) = 0.1,$
$P(S|M) = 0.8$

$P(M|S) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$

Note: posterior probability of meningitis still very small!
Other forms of Bayes Rule

\[
P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}
\]

\[
P(x \mid y) = \frac{P(y \mid x) P(x)}{\sum_x P(y \mid x) P(x)}
\]

\[
P(x \mid y) = \alpha P(y \mid x) P(x)
\]

posterior \propto \text{likelihood} \cdot \text{prior}
Conditional Bayes Rule

\[ P(x \mid y, z) = \frac{P(y \mid x, z) \ P(x \mid z)}{P(y \mid z)} \]

\[ P(x \mid y, z) = \frac{P(y \mid x, z) \ P(x, z)}{\sum_x P(y \mid x, z) \ P(x \mid z)} \]

\[ P(x \mid y, z) = \alpha P(y \mid x, z) P(x \mid z) \]
Bayes’ Rule & Cond. Independence

\[ P(\text{Cavity}|\text{toothache} \land \text{catch}) = \alpha P(\text{toothache} \land \text{catch}|\text{Cavity}) P(\text{Cavity}) \]
\[ = \alpha P(\text{toothache}|\text{Cavity}) P(\text{catch}|\text{Cavity}) P(\text{Cavity}) \]

This is an example of a \textit{naive Bayes} model:

\[ P(\text{Cause}, \text{Effect}_1, \ldots, \text{Effect}_n) = P(\text{Cause}) \prod_i P(\text{Effect}_i|\text{Cause}) \]

Total number of parameters is \textit{linear} in \( n \)
Simple Example of State Estimation

- Suppose a robot obtains measurement $z$
- What is $P(\text{doorOpen} \mid z)$?
Causal vs. Diagnostic Reasoning

- $P(open \mid z)$ is diagnostic.
- $P(z \mid open)$ is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

**count frequencies!**
Example

- \( P(z|\text{open}) = 0.6 \quad P(z|\neg\text{open}) = 0.3 \)
- \( P(\text{open}) = P(\neg\text{open}) = 0.5 \)

\[
P(\text{open} | z) = \frac{P(z | \text{open}) P(\text{open})}{P(z | \text{open}) P(\text{open}) + P(z | \neg\text{open}) P(\neg\text{open})}
\]

\[
P(\text{open} | z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} \approx 0.67
\]

- \( z \) raises the probability that the door is open.
Combining Evidence

• Suppose our robot obtains another observation $z_2$.

• How can we integrate this new information?

• More generally, how can we estimate $P(x| z_1...z_n)$?
Recursive Bayesian Updating

\[
P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x, z_1, \ldots, z_{n-1}) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}
\]

**Markov assumption:** \( z_n \) is independent of \( z_1, \ldots, z_{n-1} \) if we know \( x \).

\[
P(x \mid z_1, \ldots, z_n) = \frac{P(z_n \mid x, z_1, \ldots, z_{n-1}) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}
= \frac{P(z_n \mid x) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}
= \alpha \frac{P(z_n \mid x) P(x \mid z_1, \ldots, z_{n-1})}{P(z_n \mid z_1, \ldots, z_{n-1})}
= \alpha \prod_{i=1}^{n} P(z_i \mid x) P(x)
\]
Example: Second Measurement

- $P(z_2 | \text{open}) = 0.5 \quad P(z_2 | \neg \text{open}) = 0.6$
- $P(\text{open} | z_1) = \frac{2}{3}$

\[
P(\text{open} | z_2, z_1) = \frac{P(z_2 | \text{open}) P(\text{open} | z_1)}{P(z_2 | \text{open}) P(\text{open} | z_1) + P(z_2 | \neg \text{open}) P(\neg \text{open} | z_1)}
\]

\[
= \frac{1 \cdot \frac{2}{3}}{1 \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625
\]

- $z_2$ lowers the probability that the door is open.
These calculations seem laborious to do for each problem domain – is there a general representation scheme for probabilistic inference?

Yes - Bayesian Networks