Probability

- FOL fails for a domain due to:
  - **Laziness**: too much to list the complete set of rules, too hard to use the enormous rules that result
  - **Theoretical ignorance**: there is no complete theory for the domain
  - **Practical ignorance**: have not or cannot run all necessary tests
Probability

• Probability = a degree of belief

• Probability comes from:
  – Frequentist: experiments and statistical assessment
  – Objectivist: real aspects of the universe
  – Subjectivist: a way of characterizing an agent’s beliefs

• Decision theory = probability + utility theory
Probability

- **Prior probability**: probability in the absence of any other information

\[ P(\text{Dice} = 2) = \frac{1}{6} \]

**random variable**: Dice
**domain** = \(<1, 2, 3, 4, 5, 6>\)
**probability distribution**: \( P(\text{Dice}) = <\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}> \)
Probability

- **Conditional probability**: probability in the presence of some evidence

  \[ P(\text{Dice} = 2 \mid \text{Dice is even}) = \frac{1}{3} \]
  \[ P(\text{Dice} = 2 \mid \text{Dice is odd}) = 0 \]

- \( P(A \mid B) = \frac{P(A \cap B)}{P(B)} \)

  \[ P(A \cap B) = P(A \mid B).P(B) \]
Probability

• Example:

$S =$ stiff neck

$M =$ meningitis

$P(S \mid M) = 0.5$

$P(M) = 1/50000$

$P(S) = 1/20$

$\Rightarrow P(M \mid S) = P(S \mid M).P(M)/P(S) = 1/5000$
Probability

- Joint probability distributions:

\[ X: <x_1, \ldots, x_m> \quad Y: <y_1, \ldots, y_n> \]

\[ P(X = x_i, Y = y_j) \]
Probability

- **Axioms:**

\[
0 \leq P(A) \leq 1 \\
P(\text{true}) = 1 \quad \text{and} \quad P(\text{false}) = 0 \\
P(A \lor B) = P(A) + P(B) - P(A \land B)
\]
Probability

**Derived properties:**

\[ P(\neg A) = 1 - P(A) \]

\[ P(U) = P(A_1) + P(A_2) + \ldots + P(A_n) \]

\[ U = A_1 \lor A_2 \lor \ldots \lor A_n \text{ collectively exhaustive} \]

\[ A_i \land A_j = \text{false} \text{ mutually exclusive} \]
Probability

• Bayes’ theorem:

\[ P(H_i \mid E) = \frac{P(H_i \wedge E)}{P(E)} = \frac{P(E \mid H_i) \cdot P(H_i)}{\sum P(E \mid H_i) \cdot P(H_i)} \]

Hi’s are collectively exhaustive & mutually exclusive
Probability

- A full joint probability distribution $P(X_1, X_2, \ldots, X_n)$ is sufficient for computing any probability on $X_i$'s.
Probability

- **Problem**: a full joint probability distribution \( P(X_1, X_2, ..., X_n) \) is sufficient for computing any probability on \( X_i \)'s, but the number of joint probabilities is exponential.
Probability

• **Independence:**

\[ P(A \land B) = P(A) \cdot P(B) \]

\[ P(A) = P(A \mid B) \]

• **Conditional independence:**

\[ P(A \land B \mid E) = P(A \mid E) \cdot P(B \mid E) \]

\[ P(A \mid E) = P(A \mid E \land B) \]
Probability

- Example:

\[ P(\text{Toothache} \mid \text{Cavity} \land \text{Catch}) = P(\text{Toothache} \mid \text{Cavity}) \]

\[ P(\text{Catch} \mid \text{Cavity} \land \text{Toothache}) = P(\text{Catch} \mid \text{Cavity}) \]
Probability

"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."
Probability

"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."

Q1: If earthquake happens, how likely will John make a call?


"In John's and Mary's house, an alarm is installed to sound in case of burglary or earthquake. When the alarm sounds, John and Mary may make a call for help or rescue."

Q1: If earthquake happens, how likely will John make a call?

Q2: If the alarm sounds, how likely is the house burglarized?
Bayesian Networks


• 2000 AAAI Classic Paper Award
Bayesian Networks

- **Burglary** with $P(B) = 0.001$
- **Earthquake** with $P(E) = 0.002$
- **Alarm**
- **JohnCalls**
  - $P(J)$:
    - T: 0.90
    - F: 0.05
- **MaryCalls**
  - $P(M)$:
    - T: 0.70
    - F: 0.01

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Bayesian Networks

• **Syntax:**
  
  – A set of random variables making up the nodes
  – A set of directed links connecting pairs of nodes
  – Each node has a conditional probability table that quantifies the effects of its parent nodes
  – The graph has no directed cycles

![Diagram of Bayesian Networks with nodes A, B, C, D and arrows indicating dependencies between them, with labels yes and no indicating the absence or presence of cycles.]
Bayesian Networks

- **Semantics:**
  
  Ordering the nodes: $X_i$ is a predecessor of $X_j \Rightarrow i < j$

  \[
P(X_1, X_2, \ldots, X_n) = P(X_n | X_{n-1}, \ldots, X_1).P(X_{n-1} | X_{n-2}, \ldots, X_1). \ldots .P(X_2 | X_1).P(X_1)
  \]

  \[
  = \Pi_i P(X_i | X_{i-1}, \ldots, X_1) = \Pi_i P(X_i | \text{Parents}(X_i))
  \]

  $\text{Parents}(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$

  Each node is conditionally independent of its predecessors given its parents
Bayesian Networks

• Example:

\[
P(J \land M \land A \land \neg B \land \neg E)
= P(J \mid A).P(M \mid A).P(A \mid \neg B \land \neg E).P(\neg B).P(\neg E)
\]
Bayesian Networks

= Conditional Probabilities + Independence Assumptions

= Full Joint Probability Distribution
Bayesian Networks

- \( P(\text{Query} | \text{Evidence}) = ? \)
  - **Diagnostic** (from effects to causes): \( P(B | J) \)
  - **Causal** (from causes to effects): \( P(J | B) \)
  - **Intercausal** (between causes of a common effect): \( P(B | A, E) \)
  - **Mixed**: \( P(A | J, ¬E), P(B | J, ¬E) \)
Bayesian Networks

\[ P(B \mid A) = \frac{P(B \land A)}{P(A)} = \alpha P(B \land A) \]
Bayesian Networks

• Why Bayesian Networks?
General Conditional Independence

A node \((X)\) is conditionally independent of its non-descendents \((Z_{ij}'s)\), given its parents \((U_i's)\)
A node \((X)\) is conditionally independent of all other nodes, given its parents \((U_i's)\), children \((Y_i's)\), and children's parents \((Z_{ij}'s)\)
General Conditional Independence

X and Y are conditionally independent given E
General Conditional Independence

Example:

- Battery
  - Radio
  - Ignition
    - Starts
      - Moves
    - Gas
General Conditional Independence

Example:

Gas - (no evidence at all) - Radio
Gas ↔ Radio | Starts
Gas ↔ Radio | Moves
Network Construction

• Wrong ordering:
  – Produces a more complicated net
  – Requires more probabilities to be specified
  – Produces slight relationships that require difficult and unnatural probability judgments
  – Fails to represent all the conditional independence relationships

⇒ a lot of unnecessary probabilities
Network Construction

- **Models:**
  - **Diagnostic:** symptoms to causes
  - **Causal:** causes to symptoms

  Add "root causes" first, then the variables they influence
  ⇒ fewer and more comprehensible probabilities
Dempster-Shafer Theory

• Example: disease diagnostics

\[ D = \{ \text{Allergy, Flu, Cold, Pneumonia} \} \]
Dempster-Shafer Theory

- Example: disease diagnostics

\[ D = \{ \text{Allergy, Flu, Cold, Pneumonia} \} \]

- Basic probability assignment:

\[ m : 2^D \rightarrow [0, 1] \]

\[ m_1 : \{ \text{Flu, Cold, Pneu} \} \quad 0.6 \]

\[ D \quad 0.4 \]
Dempster-Shafer Theory

- **Belief:**

  \[ \text{Bel} : 2^D \rightarrow [0, 1] \]

  \[ \text{Bel}(S) = \sum_{X \subseteq S} m(X) \]
Dempster-Shafer Theory

• **Belief:**

\[ \text{Bel: } 2^D \rightarrow [0, 1] \]

\[ \text{Bel}(S) = \sum_{X \subseteq S} m(X) \]

• **Plausibility:**

\[ \text{Pl}(p) = 1 - \text{Bel}(\neg p) \]

\[ p: [\text{Bel}(p), \text{Pl}(p)] \quad \text{Bel}(p) + \text{Bel}(\neg p) \leq 1 \]
Dempster-Shafer Theory

- **Combination rules**: $m_1$ and $m_2$

\[
m(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y)}
\]
Exercises

• In Russell’s AIMA (2nd ed.): Chapters 13-14.