Probability

In most situations, logical deduction is not sufficient. Instead, we must make decisions based on uncertain conclusions. We can use *probabilities* to reason about uncertainty.

Let $H$ be a proposition.  

$P(H) = 1$ means that $H$ is true. 

$P(H) = 0$ means that $H$ is false. 

$P(H) = .314$ means that (we believe that) $H$ occurs 31.4% of the time.

Given a hypothesis $H$ and known evidence (facts) $E$, we would like to determine the *conditional probability* $P(H \mid E)$, the probability of $H$ given $E$. If $P(H \mid E)$ is near 1 or near 0, we can tentatively conclude $H$ or $\neg H$. Otherwise, we might try to gather more evidence $E'$ and determine $P(H \mid E, E')$, the probability of $H$ given both $E$ and $E'$.

Example: $P(\text{lab is correct} \mid \text{solved problem 1})$  

$P(\text{lab is correct} \mid \text{solved problems 1 and 2})$
Probability Agent

function PROBABILITY-AGENT()
static: states, actions
loop
  percept ← perceive environment
  current ← POSSIBLE-STATES(percept)
  for each action in actions
    next ← POSSIBLE-STATES(current, action)
    value(action) ←
      sum over $S \in$ states of $P(S \mid next) \ast value(S)$
  choose action with maximum value
  perform action on environment

Probability Theory

Probabilities satisfy the following axioms:

1. For any proposition $A$, $0.0 \leq P(A) \leq 1.0$.
2. $P(True) = 1.0$ and $P(False) = 0.0$.
3. For any two propositions $A$ and $B$,
   $P(A \lor B) = P(A) + P(B) - P(A \land B)$

Conditional probabilities are defined by:

$$P(H \mid E) = \frac{P(H \land E)}{P(E)}$$
This implies Bayes’ Theorem:

\[ P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} \]

A joint probability distribution over several propositions \( P(A_1, \ldots, A_n) \) assigns a probability to every value assignment.

The sum of the probabilities of a joint probability distribution is 1.

---

Probability Rules

Sum of Mutually Exclusive Outcomes:

1 = \( P(A) + P(\neg A) \)

1 = \( P(A \land B) + P(A \land \neg B) + P(\neg A \land B) + P(\neg A \land \neg B) \)

Marginal Distribution Rule:

\( P(A) = P(A \land B) + P(A \land \neg B) \)

\( P(A \land B) = \)

\[ P(A \land B \land C \land D) + P(A \land B \land C \land \neg D) + P(A \land B \land \neg C \land D) + P(A \land B \land \neg C \land \neg D) \]
Independence: \(A\) and \(B\) are independent iff
\[
\begin{align*}
P(A \land B) &= P(A) \cdot P(B) \\
P(A \land \neg B) &= P(A) \cdot P(\neg B) \\
P(\neg A \land B) &= P(\neg A) \cdot P(B) \\
P(\neg A \land \neg B) &= P(\neg A) \cdot P(\neg B)
\end{align*}
\]

Conditional Independence:
\(A\) and \(B\) are independent given \(C\) iff
\[
\begin{align*}
P(A \land B \mid C) &= P(A \mid C) \cdot P(B \mid C) \\
P(A \land \neg B \mid C) &= P(A \mid C) \cdot P(\neg B \mid C) \\
P(\neg A \land B \mid C) &= P(\neg A \mid C) \cdot P(B \mid C) \\
P(\neg A \land \neg B \mid C) &= P(\neg A \mid C) \cdot P(\neg B \mid C)
\end{align*}
\]

Joint Distribution Example

<table>
<thead>
<tr>
<th>(P(A, B, C, D))</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) (B) (C) (D)</td>
<td></td>
</tr>
<tr>
<td>(T) (T) (T) (T)</td>
<td>0.096</td>
</tr>
<tr>
<td>(T) (T) (T) (F)</td>
<td>0.144</td>
</tr>
<tr>
<td>(T) (T) (F) (T)</td>
<td>0.128</td>
</tr>
<tr>
<td>(T) (T) (F) (F)</td>
<td>0.032</td>
</tr>
<tr>
<td>(T) (F) (T) (T)</td>
<td>0.036</td>
</tr>
<tr>
<td>(T) (F) (T) (F)</td>
<td>0.024</td>
</tr>
<tr>
<td>(T) (F) (F) (T)</td>
<td>0.008</td>
</tr>
<tr>
<td>(T) (F) (F) (F)</td>
<td>0.032</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(P(A, B, C, D))</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) (B) (C) (D)</td>
<td></td>
</tr>
<tr>
<td>(F) (T) (T) (T)</td>
<td>0.016</td>
</tr>
<tr>
<td>(F) (T) (T) (F)</td>
<td>0.024</td>
</tr>
<tr>
<td>(F) (T) (F) (T)</td>
<td>0.128</td>
</tr>
<tr>
<td>(F) (T) (F) (F)</td>
<td>0.032</td>
</tr>
<tr>
<td>(F) (F) (T) (T)</td>
<td>0.036</td>
</tr>
<tr>
<td>(F) (F) (T) (F)</td>
<td>0.024</td>
</tr>
<tr>
<td>(F) (F) (F) (T)</td>
<td>0.048</td>
</tr>
<tr>
<td>(F) (F) (F) (F)</td>
<td>0.192</td>
</tr>
</tbody>
</table>
Combining Evidence Assuming Conditional Independence

Suppose we want to determine:

\[ P(\text{lab is correct} \mid \text{solved problems 1 thru 100}) \]

A joint probability table would be too large. Better is assuming conditional independence.

\[ P(\text{lab is correct} \mid \text{solved problems 1 thru 100}) = \alpha P(\text{lab is correct}) \prod_{i=1}^{100} P(\text{solved problem } i \mid \text{lab is correct}) \]

where \( \alpha \) is a normalization constant.

---

Bayesian Networks

A Bayesian network is an acyclic directed graph, where the nodes are variables and the edges are dependencies. If \( A \) causally influences \( B \), there should be a path from \( A \) to \( B \).

For each node \( X_i \), we need to specify how it depends on its parents:

\[ P(X_i \mid \text{Parents}(X_i)) \]

The parents of \( X_i \) should causally affect \( X_i \).
The joint distribution is specified by:

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

$X$ is independent of $Y$ given $E$ iff for all undirected paths $X, Z_1, \ldots, Z_m, Y$, there exists a $Z_i$ in the path such that:

- $Z_i \in E$ and $Z_i$ causes $Z_{i-1}$ and/or $Z_{i+1}$, or
- $Z_{i-1}$ and $Z_{i+1}$ cause $Z_i$, and none of $Z_i$ and its descendents are in $E$.

Example of a Bayesian Network

- $H =$ hardware problem
- $B =$ bug in lab
- $M =$ mailer is running
- $L =$ lab is running
- $R =$ received email
- $S =$ problem solved

Each node needs a probability table. Size of table depends on number of parents.
### Calculation for Bayesian Networks

Brute force calculation of $P(H \mid E)$ is done by:

1) Apply the conditional probability rule.

$$P(H \mid E) = \frac{P(H \land E)}{P(E)}$$

2) Apply the marginal distribution rule to the unknown vertices $U$.

$$P(H \land E) = \sum_{U=u} P(H \land E \land U = u)$$
3) Apply joint distribution rule for Bayesian networks.

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

4) (optional) Instead of using all unknowns \( U \), use only those unknowns which are dependent.

---

Example Calculation

Calculate \( P(B \mid \neg R, S) \) in the buggy lab example.

1) Apply the conditional probability rule.

\[ P(B \mid \neg R, S) = \frac{P(B, \neg R, S)}{P(\neg R, S)} \]

2) Apply the marginal distribution rule to the unknown vertices. \( P(B, \neg R, S) \) has 3 unknown vertices with \( 2^3 = 8 \) possible value assignments.
\[ P(B, \neg R, S) \]
\[ = P(B, \neg R, S, H, M, L) \]
\[ + P(B, \neg R, S, H, M, \neg L) \]
\[ + P(B, \neg R, S, H, \neg M, L) \]
\[ + P(B, \neg R, S, H, \neg M, \neg L) \]
\[ + P(B, \neg R, S, \neg H, M, L) \]
\[ + P(B, \neg R, S, \neg H, M, \neg L) \]
\[ + P(B, \neg R, S, \neg H, \neg M, L) \]
\[ + P(B, \neg R, S, \neg H, \neg M, \neg L) \]

3) Apply joint distribution rule for Bayesian networks. Here are two examples.

\[ P(B, \neg R, S, H, M, L) \]
\[ = P(B)P(H) \]
\[ \quad P(M \mid H)P(\neg R \mid M) \]
\[ \quad P(L \mid H, M)P(S \mid L) \]

\[ P(B, \neg R, S, \neg H, M, \neg L) \]
\[ = P(B)P(\neg H) \]
\[ \quad P(M \mid \neg H)P(\neg R \mid M) \]
\[ \quad P(\neg L \mid \neg H, M)P(S \mid \neg L) \]