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 $\leftarrow$ perceive environment
    current $\leftarrow$ POSSIBLE-STATES(*percept*)
    for each *action* in *actions*
      next $\leftarrow$ POSSIBLE-STATES(*current*, *action*)
      value(*action*) $\leftarrow$
        sum over $S \in states$ of $P(S \mid next) * value(S)$
    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,
$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Conditional probabilities are defined by:

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

This implies Bayes' Theorem:

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

A joint probability distribution over several propositions $\mathbf{P}(A_1, \ldots, A_n)$ 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:

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

$$\begin{aligned} 1 = {} & P(A \wedge B) + P(A \wedge \neg B) + \\ & P(\neg A \wedge B) + P(\neg A \wedge \neg B) \end{aligned}$$

Marginal Distribution Rule:

$$P(A) = P(A \wedge B) + P(A \wedge \neg B)$$

$$\begin{aligned} & P(A \wedge B) = \\ & P(A \wedge B \wedge C \wedge D) + P(A \wedge B \wedge C \wedge \neg D) + \\ & P(A \wedge B \wedge \neg C \wedge D) + P(A \wedge B \wedge \neg C \wedge \neg D) \end{aligned}$$

Independence: A and B are independent iff

$$\begin{aligned} P(A \wedge B) &= P(A)\ P(B) \\ P(A \wedge \neg B) &= P(A)\ P(\neg B) \\ P(\neg A \wedge B) &= P(\neg A)\ P(B) \\ P(\neg A \wedge \neg B) &= P(\neg A)\ P(\neg B) \end{aligned}$$

Conditional Independence:
A and B are independent given C iff

$$\begin{aligned} P(A \wedge B \mid C) &= P(A \mid C)\ P(B \mid C) \\ P(A \wedge \neg B \mid C) &= P(A \mid C)\ P(\neg B \mid C) \\ P(\neg A \wedge B \mid C) &= P(\neg A \mid C)\ P(B \mid C) \\ P(\neg A \wedge \neg B \mid C) &= P(\neg A \mid C)\ P(\neg B \mid C) \end{aligned}$$

Joint Distribution Example

$\mathbf{P}(A,B,C,D)$				
A	B	C	D	P
T	T	T	T	0.096
T	T	T	F	0.144
T	T	F	T	0.128
T	T	F	F	0.032
T	F	T	T	0.036
T	F	T	F	0.024
T	F	F	T	0.008
T	F	F	F	0.032

$\mathbf{P}(A,B,C,D)$				
A	B	C	D	P
F	T	T	T	0.016
F	T	T	F	0.024
F	T	F	T	0.128
F	T	F	F	0.032
F	F	T	T	0.036
F	F	T	F	0.024
F	F	F	T	0.048
F	F	F	F	0.192

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 α 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:

$$\mathbf{P}(X_i \mid Parents(X_i))$$

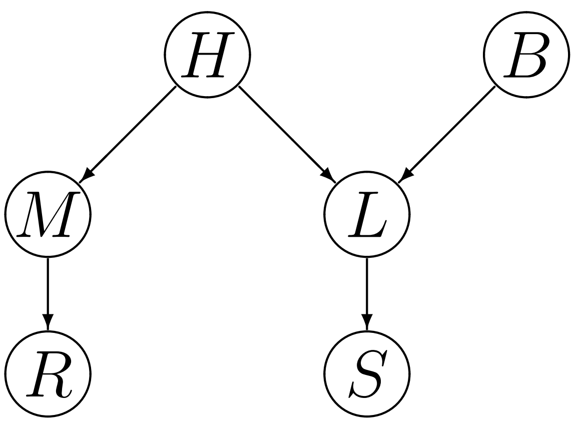
The parents of X_i should causally affect X_i.

The joint distribution is specified by:

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

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.

$\mathbf{P}(H)$	
True	*False*
0.01	0.99

H	$\mathbf{P}(M \mid H)$ *True*	*False*
True	0.1	0.9
False	0.99	0.01

H	B	$\mathbf{P}(L \mid H, B)$ *True*	*False*
True	*True*	0.01	0.99
True	*False*	0.1	0.9
False	*True*	0.02	0.98
False	*False*	1.0	0.0

Calculation for Bayesian Networks

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

1) Apply the conditional probability rule.

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

2) Apply the marginal distribution rule to the unknown vertices $\mathbf{U}$.

$$P(H \wedge E) = \sum_{\mathbf{U}=\mathbf{u}} P(H \wedge E \wedge \mathbf{U} = \mathbf{u})$$

3) Apply joint distribution rule for Bayesian networks.

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

4) (optional) Instead of using all unknowns $\mathbf{U}$, 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.

$$
\begin{aligned}
&P(B, \neg R, S) \\
&= P(B, \neg R, S, H, M, L) \\
&\quad + P(B, \neg R, S, H, M, \neg L) \\
&\quad + P(B, \neg R, S, H, \neg M, L) \\
&\quad + P(B, \neg R, S, H, \neg M, \neg L) \\
&\quad + P(B, \neg R, S, \neg H, M, L) \\
&\quad + P(B, \neg R, S, \neg H, M, \neg L) \\
&\quad + P(B, \neg R, S, \neg H, \neg M, L) \\
&\quad + P(B, \neg R, S, \neg H, \neg M, \neg L)
\end{aligned}
$$

---

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

$$
\begin{aligned}
&P(B, \neg R, S, H, M, L) \\
&= P(B)P(H) \\
&\quad P(M \mid H)P(\neg R \mid M) \\
&\quad P(L \mid H, M)P(S \mid L)
\end{aligned}
$$

$$
\begin{aligned}
&P(B, \neg R, S, \neg H, M, \neg L) \\
&= P(B)P(\neg H) \\
&\quad P(M \mid \neg H)P(\neg R \mid M) \\
&\quad P(\neg L \mid \neg H, M)P(S \mid \neg L)
\end{aligned}
$$