Probability

Contents

Probability

Motivation

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. The idea of probabilities are:

- 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 (in some sense) H has a 31.4% of being true.

Evidence and Conditional Probability

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: $P($lab is correct | solved problem $1)$
$P($lab is correct | solved problems 1 and $2)$

Probability Agent

function PROBABILITY-AGENT()
  static: *actions*
  loop
    percept ← perceive environment
    current ← POSSIBLE-STATES(*percept*)
    for each *action* in *actions*
      next ← POSSIBLE-STATES(*current*, *action*)
      E ← current evidence and proposed action
      value(*action*) ←
          sum $P(s \mid E) * value(s)$ for each $s \in next$
    choose *action* with maximum value
    perform *action* on environment

Some Probability Basics

Probabilities satisfy the following properties:

- For any proposition A, $0.0 \leq P(A) \leq 1.0$.
- $P(True) = 1.0$ and $P(False) = 0.0$.
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- $P(H \mid E) = P(H \wedge E) \,/\, P(E)$
- Bayes' Theorem: $P(H \mid E) = P(E \mid H)\, P(H) \,/\, P(E)$
- 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, Part 1

Sum of Mutually Exclusive Outcomes:

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

Marginal Distribution Rule:

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

Probability Rules, Part 2

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	0.04
T	T	T	F	0.04
T	T	F	T	0.32
T	T	F	F	0.00
T	F	T	T	0.00
T	F	T	F	0.08
T	F	F	T	0.16
T	F	F	F	0.16

$\mathbf{P}(A,B,C,D)$				
A	B	C	D	P
F	T	T	T	0.01
F	T	T	F	0.01
F	T	F	T	0.02
F	T	F	F	0.00
F	F	T	T	0.00
F	F	T	F	0.08
F	F	F	T	0.04
F	F	F	F	0.04

Combining Evidence & Conditional Independence

- Suppose we want to determine:

$$P(H \mid E_1, \ldots, E_n)$$

- A joint probability table will be too large.
- Better (but naive) is assuming conditional independence.

$$\begin{aligned} & P(H \mid E_1, \ldots, E_n) \\ & = P(H)\, P(E_1, \ldots, E_n \mid H) \,/\, P(E_1, \ldots, E_n) \\ & \approx P(H) \prod_{i=1}^{n} P(E_i|H) \,/\, P(E_1, \ldots, E_n) \end{aligned}$$

- Different hypotheses have same denominator, so only need to compare numerators.

Bayesian Networks

Definition of 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 directly affect X_i (in contrast to other variables).

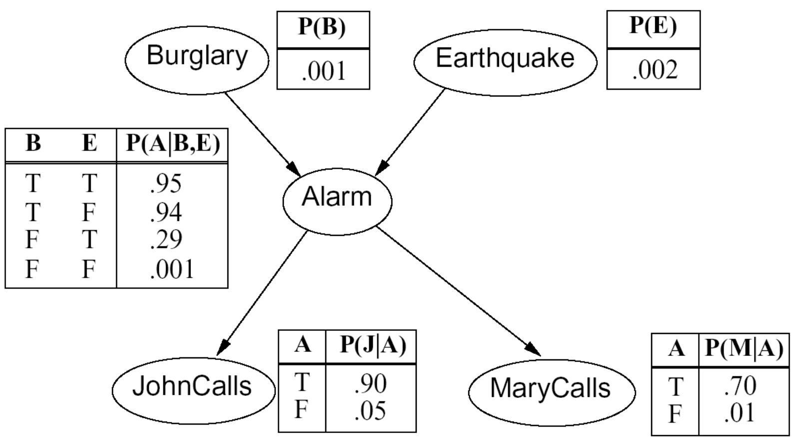
Joint Distribution

- 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 conditionally independent of Y given X's parents if Y is not a descendant of X,
- X is conditionally independent of Y given X's parents, X's children, and the parents of X's children.

Example of a Bayesian Network



P(B)
.001

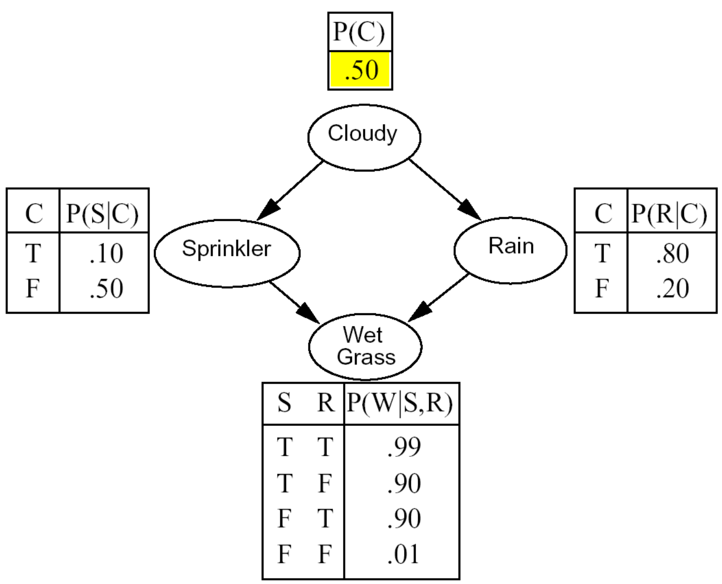
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Example of a Bayesian Network



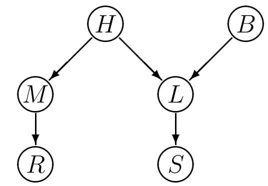
P(C)
.50

C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

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.

Examples of Conditional Probability Tables

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

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

		$\mathbf{P}(L \mid H, B)$	
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.

$$P(H \mid E) = P(H \wedge E) \,/\, P(E)$$

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.

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

Example Calculation, Part 1

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.

Example Calculation, Part 2

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

Example Calculation, Part 3

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

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

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