Uncertainty

Chapter 14

Outline

- ◇ Uncertainty
- ◇ Probability
- ◇ Syntax
- ◇ Semantics
- ◇ Inference rules

Uncertainty

Let action A_t = leave for airport t minutes before flight
Will A_t get me there on time?

Problems:
1) partial observability (road state, other drivers' plans, etc.)
2) noisy sensors (KCBS traffic reports)
3) uncertainty in action outcomes (flat tire, etc.)
4) immense complexity of modelling and predicting traffic

Hence a purely logical approach either
1) risks falsehood: "A_{25} will get me there on time"
or 2) leads to conclusions that are too weak for decision making:
"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc etc."

(A_{1440} might reasonably be said to get me there on time
but I'd have to stay overnight in the airport ...)

Methods for handling uncertainty

Default or nonmonotonic logic:
Assume my car does not have a flat tire
Assume A_{25} works unless contradicted by evidence
Issues: What assumptions are reasonable? How to handle contradiction?

Rules with fudge factors:
$A_{25} \mapsto_{0.3}$ get there on time
$Sprinkler \mapsto_{0.99} WetGrass$
$WetGrass \mapsto_{0.7} Rain$
Issues: Problems with combination, e.g., $Sprinkler$ causes $Rain$??

Probability
Given the available evidence,
A_{25} will get me there on time with probability 0.04
Mahaviracarya (9th C.), Cardamo (1565) theory of gambling

(Fuzzy logic handles *degree of truth* NOT uncertainty e.g.,
$WetGrass$ is true to degree 0.2)

Probability

Probabilistic assertions *summarize* effects of
laziness: failure to enumerate exceptions, qualifications, etc.
ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:
Probabilities relate propositions to one's own state of knowledge
e.g., $P(A_{25}|\text{no reported accidents}) = 0.06$

These are not assertions about the world

Probabilities of propositions change with new evidence:
e.g., $P(A_{25}|\text{no reported accidents, 5 a.m.}) = 0.15$

(Analogous to logical entailment status $KB \models \alpha$, not truth.)

Making decisions under uncertainty

Suppose I believe the following:

$$P(A_{25}\text{ gets me there on time}|\ldots) = 0.04$$
$$P(A_{90}\text{ gets me there on time}|\ldots) = 0.70$$
$$P(A_{120}\text{ gets me there on time}|\ldots) = 0.95$$
$$P(A_{1440}\text{ gets me there on time}|\ldots) = 0.9999$$

Which action to choose?

Depends on my preferences for missing flight vs. airport cuisine, etc.

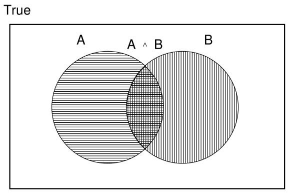
Utility theory is used to represent and infer preferences

Decision theory = utility theory + probability theory

Axioms of probability

For any propositions A, B

1. $0 \leq P(A) \leq 1$
2. $P(True) = 1$ and $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Syntax

Similar to propositional logic: possible worlds defined by assignment of values to random variables.

Propositional or Boolean random variables
e.g., $Cavity$ (do I have a cavity?)
Include propositional logic expressions
e.g., $\neg Burglary \vee Earthquake$

Multivalued random variables
e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow\rangle$
Values must be exhaustive and mutually exclusive

Proposition constructed by assignment of a value:
e.g., $Weather = sunny$; also $Cavity = true$ for clarity

Syntax contd.

Prior or unconditional probabilities of propositions
e.g., $P(Cavity) = 0.1$ and $P(Weather = sunny) = 0.72$
correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:
$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1\rangle$ (normalized, i.e., sums to 1)

Joint probability distribution for a set of variables gives values for each possible assignment to all the variables
$\mathbf{P}(Weather, Cavity)$ = a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$				
$Cavity = false$				

Syntax contd.

Conditional or posterior probabilities
e.g., $P(Cavity|Toothache) = 0.8$
i.e., given that $Toothache$ is all I know

Notation for conditional distributions:
$\mathbf{P}(Weather|Earthquake)$ = 2-element vector of 4-element vectors

If we know more, e.g., $Cavity$ is also given, then we have
$P(Cavity|Toothache, Cavity) = 1$
Note: the less specific belief *remains valid* after more evidence arrives, but is not always *useful*

New evidence may be irrelevant, allowing simplification, e.g.,
$P(Cavity|Toothache, 49ersWin) = P(Cavity|Toothache) = 0.8$
This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

Definition of conditional probability:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \text{ if } P(B) \neq 0$$

Product rule gives an alternative formulation:
$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$

A general version holds for whole distributions, e.g.,
$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$
(View as a 4×2 set of equations, *not* matrix mult.)

Chain rule is derived by successive application of product rule:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n_1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \Pi_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Bayes' Rule

Product rule $P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$

$$\Rightarrow \text{ Bayes' rule } P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Why is this useful???

For assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g., let M be meningitis, S be stiff neck:

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

Note: posterior probability of meningitis still very small!

Normalization

Suppose we wish to compute a posterior distribution over A given $B=b$, and suppose A has possible values $a_1 \ldots a_m$

We can apply Bayes' rule for each value of A:

$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$

$\ldots$

$P(A=a_m|B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$

Adding these up, and noting that $\Sigma_i P(A=a_i|B=b) = 1$:

$$1/P(B=b) = 1/\Sigma_i P(B=b|A=a_i)P(A=a_i)$$

This is the normalization factor, constant w.r.t. i, denoted α:

$$\mathbf{P}(A|B=b) = \alpha \mathbf{P}(B=b|A)\mathbf{P}(A)$$

Typically compute an unnormalized distribution, normalize at end

e.g., suppose $\mathbf{P}(B=b|A)\mathbf{P}(A) = \langle 0.4, 0.2, 0.2 \rangle$

then $\mathbf{P}(A|B=b) = \alpha\langle 0.4, 0.2, 0.2 \rangle = \frac{\langle 0.4,0.2,0.2 \rangle}{0.4+0.2+0.2} = \langle 0.5, 0.25, 0.25 \rangle$

Conditioning

Introducing a variable as an extra condition:

$$P(X|Y) = \Sigma_z P(X|Y, Z=z)P(Z=z|Y)$$

Intuition: often easier to assess each specific circumstance, e.g.,

$P(RunOver|Cross)$

$= P(RunOver|Cross, Light=green)P(Light=green|Cross)$

$+ P(RunOver|Cross, Light=yellow)P(Light=yellow|Cross)$

$+ P(RunOver|Cross, Light=red)P(Light=red|Cross)$

When Y is absent, we have summing out or marginalization:

$$P(X) = \Sigma_z P(X|Z=z)P(Z=z) = \Sigma_z P(X, Z=z)$$

In general, given a joint distribution over a set of variables, the distribution over any subset (called a marginal distribution for historical reasons) can be calculated by summing out the other variables.

Full joint distributions

A complete probability model specifies every entry in the joint distribution for all the variables $\mathbf{X} = X_1, \ldots, X_n$

I.e., a probability for each possible world $X_1=x_1, \ldots, X_n=x_n$

(Cf. complete theories in logic.)

E.g., suppose $Toothache$ and $Cavity$ are the random variables:

	$Toothache=true$	$Toothache=false$
$Cavity=true$	0.04	0.06
$Cavity=false$	0.01	0.89

Possible worlds are mutually exclusive $\Rightarrow$ $P(w_1 \wedge w_2) = 0$

Possible worlds are exhaustive $\Rightarrow$ $w_1 \vee \cdots \vee w_n$ is $True$

hence $\Sigma_i P(w_i) = 1$

Full joint distributions contd.

1) For any proposition ϕ defined on the random variables

$\phi(w_i)$ is true or false

2) ϕ is equivalent to the disjunction of w_is where $\phi(w_i)$ is true

Hence $P(\phi) = \Sigma_{\{w_i:\ \phi(w_i)\}} P(w_i)$

I.e., the unconditional probability of any proposition is computable as the sum of entries from the full joint distribution

Conditional probabilities can be computed in the same way as a ratio:

$$P(\phi|\xi) = \frac{P(\phi \wedge \xi)}{P(\xi)}$$

E.g.,

$$P(Cavity|Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = \frac{0.04}{0.04 + 0.01} = 0.8$$

Inference from joint distributions

Typically, we are interested in

the posterior joint distribution of the query variables $\mathbf{Y}$

given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:

1) Worst-case time complexity $O(d^n)$ where d is the largest arity

2) Space complexity $O(d^n)$ to store the joint distribution

3) How to find the numbers for $O(d^n)$ entries???