Uncertainty

AIMA2e Chapter 13

Outline

- ♢ Uncertainty
- ♢ Probability
- ♢ Syntax and Semantics
- ♢ Inference
- ♢ Independence and Bayes' Rule

FOL and Uncertainty

One problem with FOL is that it can not handle uncertainty:

Assume two propositions:

A=Leaving for the airport 2hrs before your flight
B=Catching the flight

Then, you can *reasonably* say:

$$A \Rightarrow B$$

But this is not always True.

Uncertainty

Uncertainty is due to either:

◊ Laziness: too much work to list all the antecedents to ensure an exceptionless rule:

$$A \wedge \neg Rain \wedge \neg Traffic \ldots \;\Rightarrow\; B$$

◊ Ignorance: We may not have a complete theory for the domain (e.g. medical)

Even if we listed a long precedent list, we may not be able to apply such a rule because we may have only partial information

Uncertainty in FOL

Let action $A_t =$ leave for airport t minutes before flight

A purely logical approach either
1) risks falsehood: "A_{30} will get me there on time" or
2) leads to conclusions that are too weak for decision making:
    "A_{30} 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."
3) be overly pessimistic:
    A_{1440} will get me there on time but I'd have to stay overnight in the airport . . .)

Methods for handling uncertainty

Making assumptions:

Assume my car does not have a flat tire

Assume there is no outrageous traffic...

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 $Rain$??

Fuzzy logic:

Handles *degree of truth* NOT uncertainty e.g.,

$WetGrass$ is true to degree 0.2

Probability(Mahaviracarya(9th C.), Cardamo(1565) - theory of gambling):

Given the available evidence,

A_{25} will get me there on time with probability 0.04

Probability

- Probabilities relate propositions to one's own state of knowledge, by assigning a numerical <u>degree of belief</u> between 0 and 1, to each event.
  $P(A_{25}$ gets me there on time$|$no reported accidents$) = 0.06$
- Probabilities of propositions change with new evidence:
  $P(A_{25}$ gets me there on time$|$no reported accidents, 5 a.m.$) = 0.15$

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

Making decisions under uncertainty

Suppose then that I believe the following:

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

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

Probability basics

◊ Begin with a set Ω—the *sample space*
e.g., 6 possible rolls of a die.
$\omega \in \Omega$ is a sample point/possible world/atomic event

◊ A *probability space* or *probability model* is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.
$0 \leq P(\omega) \leq 1$
$\sum_{\omega} P(\omega) = 1$
e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

◊ An *event* A is any subset of Ω

$$P(A) = \sum_{\{\omega \in A\}} P(\omega)$$

E.g., $P(\text{die roll} < 4) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables and Probability

A *random variable* is a function that maps outcomes of some random experiment to some range, e.g., Real or Boolean values

♢ A random variable $DiceValue$, can be used to describe the outcome of rolling a fair die to the possible outcomes 1, 2, 3, 4, 5, 6 .

♢ A random variable $OddDice$, can be used to describe the evenness-oddness of the rolled dice with possible outcomes True,False.

♢ P induces a *probability distribution* for any random variable X:

$$P(X = x_i) = \Sigma_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

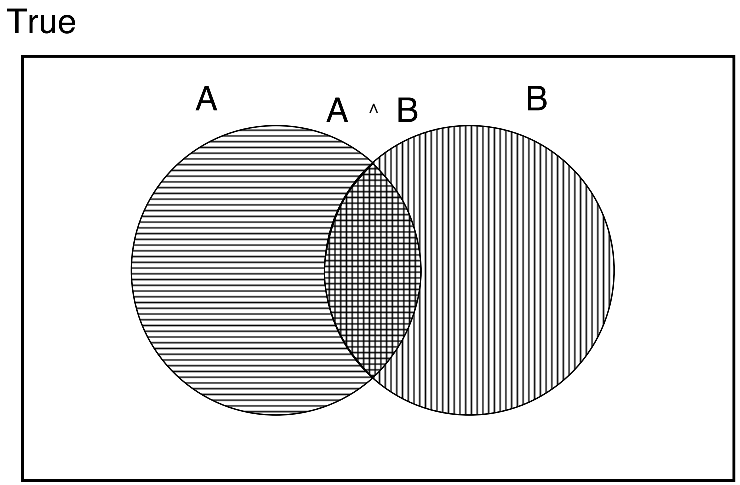
where $\omega \in \Omega$ is the samples in the *sample space*

e.g., $P(OddDice = true) = 1/6 + 1/6 + 1/6 = 1/2$
i.e., the event is the random variable taking certain values.

Why use probability?

The definitions imply that certain logically related events must have related probabilities

E.g., $P(a \lor b) = P(a) + P(b) - P(a \land b)$



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.

Probability of Propositions

We will talk about probabilities of propositions composed of statements about random variable outcomes:

◇ Propositional or Boolean random variables
e.g., $Cavity$ (do I have a cavity?)

◇ Discrete random variables (*finite* or *infinite*)
e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$
Values must be exhaustive and mutually exclusive

◇ Continuous random variables (*bounded* or *unbounded*)
e.g., $Temp \in \Re$

◇ Arbitrary Boolean combinations of basic propositions
$Weather = rain$ is a proposition
$Temp < 22.0$ is a proposition

Prior probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.1$ 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 r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Joint Distribution

Joint probability distribution for a set of variables gives values **for each possible assignment to all the variables**

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

Note that the entries sum up to 1 (mutually exclusive and exhaustive)

Joint Distribution

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

Adding accross rows or columns give unconditional probability of a variable:

$P(Cavity) =$

$P(Cavity \vee Toothache) =$

$P(Cavity \wedge Toothache) =$

Example – In Class

Weight\Height	Short	Medium	Tall
Low	10	15	5
Medium	8	25	10
Heavy	5	10	12

N=100 people with weight and heights given as above.

- P (Weight= Low, Height=Tall) = ...
- P (Weight= Low | Height=Tall) = ...
- P (Weight= M) = ...

Joint Distribution

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

Adding accross rows or columns give unconditional probability of a variable:

$P(Cavity) = 0.04 + 0.06 = 0.10$

$P(Cavity \lor Toothache) = 0.04 + 0.06 + 0.01 = 0.11$

Alternatively: $1 - P(\neg Cavity \land \neg Toothache) = 1 - 0.89$

$P(Cavity \land Toothache) = 0.04$

Inference by enumeration

Start with the joint distribution:

	$toothache$		$\neg\, toothache$	
	$catch$	$\neg\, catch$	$catch$	$\neg\, catch$
$cavity$	**.108**	**.012**	**.072**	**.008**
$\neg\, cavity$	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$$

$$P(cavity \vee toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Conditional probability

Definition of conditional (or posterior) probability:

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

e.g., $P(Cavity|Toothache) = 0.8$

A general version holds for whole distributions, e.g.,

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

Conditional probability

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.,
$P(Cavity|Toothache, GSWon) = P(Cavity|Toothache) = 0.8$
This kind of inference, sanctioned by domain knowledge, is crucial

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

Using this formula for classification problems, we get

P(C| X) = P (X |C) P(C) / P(X)

posterior probability = α x class conditional probability x prior

Bayes' Rule

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

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

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!

Chain Rule

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

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

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:

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

Inference by enumeration

Back to our example:

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

.

We can compute conditional probabilities from joint probabilities:

$$P(Cavity|Toothache) = ?$$

Inference by enumeration

	$Toothache$	$\neg Toothache$
$Cavity$	0.04	0.06
$\neg Cavity$	0.01	0.89

.

We can compute conditional probabilities from joint probabilities:

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

Inference by enumeration - 3 variables

Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

Suppose we wish to compute the posterior probability of Meningitis given StiffNeck.

$P(m|S) = P(s|m)P(m)/P(s)$

If we are also entertaining the possibility that the patient may be suffering from a Whiplash, so we consider:

$P(w|S) = P(s|w)P(w)/P(s)$

◊ Now we can use the relative likelihoods of Meningitis and Whiplash, *without* computing $P(s)$ (divide the above two equations two cancel out $P(s)$), to find which is the more likely cause.

◊ If we still want to be able to compute $P(M|S)$, we need **normalization**...

Normalization

Consider:

$P(m|S) = P(s|m)P(m)/P(s)$ and $P(\neg m|S) = P(s|\neg m)P(\neg m)/P(s)$

Since the above two terms sum to 1 (there are only two possibilities given s), we obtain:

$P(M|s) + P(\neg m|s) = P(s|m)P(m)/P(s) + P(s|\neg m)P(\neg m) = 1$

From which we can find that:

$P(s|m)P(m) + P(s|\neg m)P(\neg m) = P(s)$

So once we have:

$$\mathbf{P}(M|s) = \langle P(s|m)P(m)/P(s), P(s|\neg m)P(\neg m)/P(s)\rangle$$

$$\mathbf{P}(M|s) = \alpha\langle P(s|m)P(m), P(s|\neg m)P(\neg m)\rangle$$

we can sum the terms of the vector ($P(s|m)P(m)$ and $P(s|\neg m)P(\neg m)$) to find α using the above equality and normalize the terms by that.

Normalization - Example

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

Denominator can be viewed as a *normalization constant*
$\alpha = 1/P(toothache)$ for this case:

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$
$$= \alpha\,(\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch))$$
$$= \alpha\,(\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle)$$
$$= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

Note: Vector indicators $\langle$ and $\rangle$ and proper Uppercase/lowercase usage, for random variable (Cavity) or given fixed evidence (toothache).

General idea: compute distribution on query variable
by fixing evidence variables and summing over hidden variables