Uncertainty

Uncertainty

What shall an agent do when not all is crystal clear?

Different types of uncertainty effecting an agent:

- The state of the world?
- The effect of actions?

Uncertain knowledge of the world:

- Inputs missing
- Limited precision in the sensors
- Incorrect model: action $\Rightarrow$ state   due to the complexity
- A changing world

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 (traffic reports)
3) uncertainty in action outcomes (flat tire, etc.)
4) immense complexity of modeling and predicting traffic

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

Right thing to do-***the rational decision***-therefore depends on both the relative importance of various goals & the likelihood that & degree to which, they will achieved.

Rational Decisions

A rational decision must consider:

- The relative importance of the sub-goals
- Utility theory
- The degree of belief that the sub-goals will be achieved
- Probability theory

Decision theory = probability theory + utility theory :

Principle of maximum expected utility (MEU)-

"The agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all possible outcomes of the action"

Using FOL for (Medical) Diagnosis

$$\forall p \;\; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$$

Not correct...

$$\forall p \;\; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity) \lor Disease(p, GumDisease) \lor Disease(p, WisdomTooth)$$

Not complete...

$$\forall p \; Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$$

Not correct...

Handling Uncertain Knowledge

Problems using first-order logic for diagnosis:

Laziness:
Too much work to make complete rules.
Too much work to use them

Theoretical ignorance:
Complete theories are rare

Practical ignorance:
We can't run all tests anyway

Probability can be used to *summarize* the laziness and ignorance !

Probability

Compare the following:

1) First-order logic:

   "The patient has a cavity"

2) Probabilistic:

   "The probability that the patient has a cavity is 0.8"

1) Is either valid or not, depending on the state of the world
2) Validity depends on the agents perception history, the evidence

Probability

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

These are **not** claims of some **probabilistic tendency** in the current situation (but might be learned from past experience of similar situations)

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

Probability

Probabilities are either:

Prior probability (unconditional , “obetingad”)

Before any evidence is obtained

Posterior probability (conditional , “betingad”)

After evidence is obtained

Probability

Notation for unconditional probability for a proposition A: P(A)

Ex: P(Cavity)=0.2 means:
"the degree of belief for "Cavity" given no extra evidence is 0.2"

Axioms for probabilities:

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

Random variable

- A random variable has a *domain* of possible values
- Each value has a assigned probability between 0 and 1
- The values are :

  Mutually exclusive (disjoint): (only one of them are true)

  Complete (there is always one that is true)

Example: The random variable Weather:

P(Weather=Sunny) = 0.7

P(Weather=Rain) = 0.2

P(Weather=Cloudy) = 0.08

P(Weather=Snow) = 0.02

Random Variable

The random variable Weather as a whole is said to have a probability distribution which is a vector (in the discrete case):

P(Weather) = [0.7 0.2 0.08 0.02]

(Notice the bold **P** which is used to denote the prob.distribution)

Random variable - Example

Example - The random variable Season:

- P(Season = Spring) = 0.26 or shorter: P(Spring)=0.26
- P(Season = Summer) = 0.20
- P(Season = Autumn) = 0.28
- P(Season = Winter) = 0.26

The random variable Season has a domain

<Spring, Summer, Autumn, Winter>

and a probability distribution:

P(Season) = [0.26 0.20 0.38 0.26]

The values in the domain are :

- Mutually exclusive (disjoint): (only one of them are true)
- Complete (there is always one that is true)

Probability Model

- 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 $\omega \in \Omega$

$$0 \le P(\omega) \le 1$$

$$\Sigma_{\omega} P(\omega) = 1$$

  e.g., $P(1) \le P(2) \le P(3) \le P(4) \le P(5) \le P(6) \le 1/6.$

- An event A is any subset of Ω

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

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

The Joint Probability Distribution

Assume that an agent describing the world using the random variables $X_1, X_2, ..X_n$.

The joint probability distribution (or ”joint”) assigns values for all combinations of values on $X_1, X_2, ..X_n$.

Notation: $\mathbf{P}(X_1, X_2, ..X_n)$ (i.e. **P** bold)

The Joint Probability Distribution

- 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)

P(Weather, Cavity) = a 4 x 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

Example - P(Season, Weather)

	Weather					
Season	Sun	Rain	Cloud	Snow		
Spring	0.07	0.03	0.10	0.06	**0.26**	P(Spring)
Summer	0.13	0.01	0.05	0.01	+**0.20**	P(Summer)
Autumn	0.05	0.05	0.15	0.03	+**0.28**	P(Autumn)
Winter	0.05	0.01	0.10	0.10	+**0.26**	P(Winter)
	0.30	**0.10**	**0.40**	**0.20**		

P(Sun)+P(Rain)+P(Cloud)+P(Snow) = **1.00**

Example: $P(Weather=Sun \wedge Season=Summer) = 0.13$

Conditional Probability

- The Posterior prob. (conditional prob.) after obtaining evidence:

Notation:

P(A|B) means: “The probability of A given that all we know is B”.

Example:

P(Sunny | Summer) = 0.65

Is defined as:

$$\boxed{P(A|B)=\frac{P(A\wedge B)}{P(B)}}$$

if $P(B) \neq 0$

Can be rewritten as the product rule:

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

“For A and B to be true, B has to be true, and A has to be true given B”

Conditional Probability

For the entire random variables:

$$\boxed{\mathbf{P}(A, B) = \mathbf{P}(B)\mathbf{P}(A \mid B)}$$

should be interpreted as a set of equations for all possible values on the random variables A and B.

Example:

$$\boxed{\mathbf{P}(Weather, Season) = \mathbf{P}(Season)\mathbf{P}(Weather \mid Season)}$$

Conditional Probability

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

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

(View as a 4 x 2 set of equations)

Chain rule is derived by successive application of product rule:

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

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:

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

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

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

Inference by enumeration

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

Can also compute conditional probabilities:

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

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

Normalization

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

Denominator can be viewed as a normalization constant α

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

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

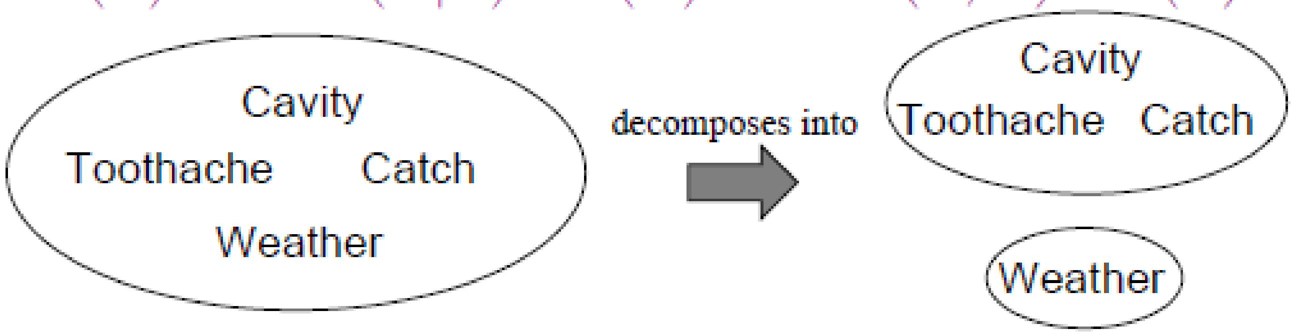
Try Yourself

- P(Toothache)
- P(Cavity)
- P(Toothache|cavity)
- P(Cavity|toothache v catch)

Independence

A and B are independent iff

$\mathbf{P}(A|B)=\mathbf{P}(A)$ or $\mathbf{P}(B|A)=\mathbf{P}(B)$ or $\mathbf{P}(A,B)=\mathbf{P}(A)\mathbf{P}(B)$



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

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

Conditional independence contd.

Write out full joint distribution using chain rule:

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

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

The left side of the product rule is symmetric w.r.t B and A:

$$P(A \wedge B) = P(A)P(B \mid A)$$
$$P(A \wedge B) = P(B)P(A \mid B)$$

Equating the two right-hand sides yields Bayes' rule:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Bayes' Rule

Useful for assessing diagnostic probability from causal probability:

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

Example of Medical Diagnosis using Bayes' rule

Known facts:

Meningitis causes stiff neck 50% of the time.

The probability of a patient having meningitis (M) is 1/50,000.

The probability of a patient having stiff neck (S) is 1/20.

Question:

What is the probability of meningitis given stiff neck ?

Solution:

P(S|M)=0.5

P(M) = 1/50,000

P(S) = 1/20

$$P(M|S) = \frac{P(S|M)\ \ P(M)}{P(S)} = \frac{0.5 \bullet 1/50000}{1/20} = 0.0002$$

Note: posterior probability of meningitis still very small!!

Example

- A doctor knows that the disease meningitis causes the patient to have a stiff neck, say, 50% of the time and whiplash 80% of the time The doctor also knows some unconditional facts: the prior probability of a patient having meningitis is 1/50,000, and having whiplash is 1/10,000. If a patient comes to a doctor which treatment will be given to him and why?

Solution

- Possibility that the patient is suffering from meningitis *W* given a stiff neck: $$P(M|S) = \frac{P(S|M) \quad P(M)}{P(S)}$$
- Possibility that the patient is suffering from whiplash *W* given a stiff neck: $$P(W|S) = \frac{P(S|W) \quad P(W)}{P(S)}$$
- Given that, *P(S\W)* = 0.8 and *P(W)* = 1/1000.

$$\frac{P(M|S)}{P(W|S)} = \frac{P(S|M)P(M)}{P(S|W)P(W)} = \frac{.5 \times \frac{1}{50000}}{.8 \times \frac{1}{10000}} = \frac{1}{80}$$

That is, whiplash is 80 times more likely than meningitis, given a stiff neck.

Bayes' Rule

In distribution form

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\,\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha\mathbf{P}(X|Y)\,\mathbf{P}(Y)$$

Combining Evidence

Task: Compute $P(Cavity|Toothache \wedge Catch)$

1. Rewrite using the definition and use the joint. With N evidence variables, the “joint” will be an N dimensional table. It is often impossible to compute probabilities for all entries in the table.

2. Rewrite using Bayes’ rule. This also requires a lot of cond.prob. to be estimated. Other methods are to prefer.

Bayes' Rule and Conditional Independence

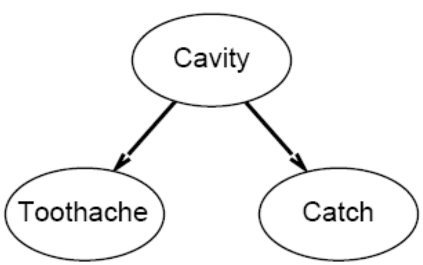
$\mathbf{P}(Cavity|toothache \wedge catch)$

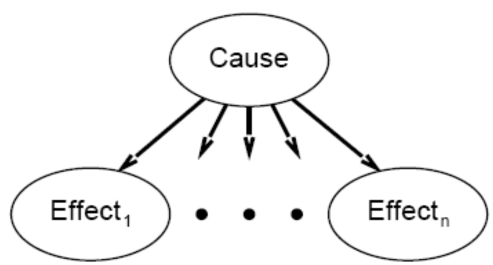
$= \alpha\ \mathbf{P}(toothache \wedge catch|Cavity)\ \mathbf{P}(Cavity)$

$= \alpha\ \mathbf{P}(toothache|Cavity)\ \mathbf{P}(catch|Cavity)\ \mathbf{P}(Cavity)$

This is an example of a **naive Bayes model**:

$\mathbf{P}(Cause,Effect_1,\ldots,Effect_n) = \mathbf{P}(Cause)\ \prod_i \mathbf{P}(Effect_i|Cause)$





Total number of parameters is linear in n

Try Yourself

- A doctor knows that pneumonia causes a fever 95% of the time. She knows that if a person is selected randomly from the population, there is a 10−7 chance of the person having pneumonia. 1 in 100 people suffer from fever. You go to the doctor complaining about the symptom of having a fever (evidence). What is the probability that pneumonia is the cause of this symptom (hypothesis)?