Uncertainty

Based on the following books-
Artificial Intelligence: Foundations of Computational Agents by D. Poole and A. Mackworth
Artificial Intelligence: A Modern Approach by S. Russell and P.Norvig

1 What is uncertainty?

We do not have complete knowledge about the world around us. But we still need to make decisions based on everything like - samples, the world, the responses, the values, signal noise etc.; which are all uncertain components.

Assuming this uncertainty isn't enough to handle it well.
For example: safety of road, security in dangerous places (think iron man), stock market stability etc.

When we take an action under uncertainty, we are gambling.
For example:
Let action A_{t} = leave for airport t minutes before flight. Will A_{t} get me there on time?

The potential unobserved variables in this situation are:

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

And a few possible scenarios are:

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.

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

A purely deterministic (a method that does not handle or model uncertainty) 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, like A_{1440}

Hence gambling with as much awareness as possible is important.

So, what do we do? We reason about it explicitly. One way to do so is by using probabilities.

2 Probability

2.1 What is probability?

Probability is an agent's measure of belief in some proposition. This is termed as *subjective probability.*

Example: Your probability of a bird flying is your measure of belief in the flying ability of an individual based only on the knowledge that the individual is a bird. This can be a potential problem because:

- Other people may have different probabilities, as they may have had different experiences with birds or different knowledge about this particular bird.
- A person's belief in a bird's flying ability is a affected by what the person knows about that bird

2.2 Use of probabilities

Probabilistic assertions summarize effects of:

- laziness: failure to enumerate exceptions, identify correct samples, etc.
- ignorance: lack of relevant facts, initial conditions, etc.

2.3 Subjective probability

Subjective probabilities relate propositions to our own state of knowledge. They change with our change of knowledge and hence are *NOT* assertions about the world.

Example:

$$\mathrm{P}(A_{25} \mid \text{no reported accidents}) = 0.06$$

This is read as probability of A_{25} *given* no reported accidents.

Probabilities of propositions change with new evidence.

Example:

$$\mathrm{P}(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$$

3 Making decisions under uncertainty

Suppose we believe the following:

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

Which amongst these should we choose?

This choice varies for each of us, depending on our *preferences* for missing flight vs. time spent waiting, etc.

These *preferences* are modeled using *Utility theory*, which is used to represent and infer preferences.

And then decisions are made using *Decision theory*, where:

Decision theory = Probability theory + Utility theory

3.1 Numerical measures of Belief

Belief in proposition, f , can be measured in terms of a number between 0 and 1, i.e., the probability of the proposition. Here:

0 - f is believed to be definitely *false*
1 - f is believed to be definitely *true*

Also, f having a probability between 0 and 1, doesn't mean f is true to some degree, but it means we are ignorant of its truth value. Probability is a measure of our ignorance, of how uncertain we are about a proposition and not how confident of it's occurrence we are.

4 Random Variables

A random variable is a term in a language that can take one of a number of different values. Now let X be a random variable, using which we will define further concepts, as follows:

- The *domain* of a variable X, written *dom(X)*, is the set of values X can take.
- A *tuple* of random variables $<X_1, ..., X_n>$is a *complex random variable* with domain $<dom(X_1), ..., dom(X_n)>$. Often the tuple is written as $X_1, ..., X_n$.
- *Assignment* X = x means variable X has value x.

- A *proposition* is a Boolean formula made from assignments of values to variables.

4.1 Important properties of a random variable

- *Elementary propositions* are of two types:

  1. Boolean random variables
     Example: *Cavity* - do I have a cavity?
  2. Discrete random variables
     Example: *Weather* is one of <sunny, rainy, cloudy, snow >

- Domain values must be mutually exhaustive and exclusive.

- Elementary proposition is constructed by assignment of a value to a random variable.

  Example:

  $Weather = sunny$

  $Cavity = false$. Sometimes abbreviated as $\neg Cavity$

- *Complex propositions* are formed from Elementary propositions and standard logical connectives.

  Example:

  $Weather = sunny \wedge Cavity = false$

- *Atomic event*: A *complete* specification of the state of the world about which the agent is uncertain.

  Example: If the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

Also, Atomic events are mutually exclusive and exhaustive.

5 Axioms of probability

For any two propositions A, B:

- $0 \leq \mathrm{P}(A) \leq 1$
- $\mathrm{P}(true) = 1$ and $\mathrm{P}(false) = 0$
- $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B)$ - $\mathrm{P}(A \wedge B)$

These axioms are sound and complete with respect to the semantics.

6 Prior probability

Prior or unconditional probabilities of propositions correspond to belief prior to arrival of any (new) evidence.

Example:
$\mathrm{P}(Cavity = true) = 0.2$ and $\mathrm{P}(Weather = sunny) = 0.72$

A *Probability distribution* gives values for all possible assignments.

Example:
$\mathrm{P}(Weather = <0.72, 0.1, 0.08, 0.1>$(*normalized*, i.e., sums to 1)

A *Joint probability distribution* for a set of random variables gives the probability of every atomic event on those random variables.

Example: $\mathrm{P}(Weather, Cavity)$ is a 4*2 matrix of values as shown below

Weather	sunny	rainy	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 this joint distribution.

7 Conditioning

Probabilistic conditioning specifies how to revise beliefs based on new information.

A probabilistic model is built taking all background information into account. This background information constitutes the prior probability. All other conclusions must be conditioned on these prior probabilities, and are called as *Conditional Probabilities.*

If evidence e is all the information obtained, subsequently, the conditional probability $\mathrm{P}(h \mid e)$, i.e., of h given e is the *Posterior Probability* of h.

7.1 Properties of Conditional Probability

- Definition:
  P(a | b) = P($a \wedge b$)/P(b) if P(b)>0
- *Product rule* gives an alternative formulation of:
  P($a \wedge b$) = P(a | b) * P(b) = P(b | a) * P(a)

7.2 Some points to note about Conditional Probability

- If toothache is all we know:
  P(*Cavity* | *Toothache*) = 0.8
- If we know more, i.e., if *Cavity* is also given:
  P(*Cavity* | *Toothache*, *Cavity*) = 1
- Notation for Conditional Distributions:
  P(*Cavity* | *Toothache*) ⇒ 2-element vector of 2-element vectors
- New evidence might be irrelevant, allowing for simplification:
  P(*Cavity* | *Toothache*, *Weather* = *sunny*) = P(*Cavity* | *Toothache*) = 0.8
  This kind of inference, sanctioned by some idea in the domain, is crucial.

7.3 Chain Rule

P($f_1 \wedge f_2 \wedge ... \wedge f_n$)
= P(f_n | $f_1 \wedge ... \wedge f_{n-1}$) * P($f_1 \wedge ... \wedge f_{n-1}$)
= P(f_n | $f_1 \wedge ... \wedge f_{n-1}$) * P($f_{n-1}$ | $f_1 \wedge ... \wedge f_{n-2}$) * P($f_1 \wedge ... \wedge f_{n-2}$)
= P(f_n | $f_1 \wedge ... \wedge f_{n-1}$) * P($f_{n-1}$ | $f_1 \wedge ... \wedge f_{n-2}$) * ... * P($f_3$ | $f_1 \wedge f_2$) * P(f_2 | f_1) * P(f_1)
= $\prod_{i=1}^{n} P(f_i|f_1 \wedge f_2 ... \wedge f_{i-1})$

8 Bayes' Theorem

The chain rule and commutativity of conjunction gives:
P($a \wedge b$) = P(a | b) * P(b) = P(b | a) * P(a)

If P(a) $\neq$ 0, we can divide both sides by P(a), giving us:

$$P(b|a) = \frac{P(a|b)*P(b)}{P(a)}$$

This is called as *Bayes' Rule* or *Bayes' Theorem*.

8.1 Why Bayes' Theorem?

We often have causal knowledge:

- P(*symptom* | *disease*)
- P(*light is off* | *state of switch*)
- P(*alarm* | *fire*)

And we want to do evidential reasoning like:

- P(*disease* | *symptom*)
- P(*state of switch* | *light is off*)
- P(*fire* | *alarm*)

Hence, Bayes' theorem is appropriately applicable, i.e., for assessing *diagnostic* probability from *causal* probability, i.e.,:

$$\mathrm{P}(Cause \mid Effect) = \mathrm{P}(Effect \mid Cause) * \mathrm{P}(Cause) \;/\; \mathrm{P}(Effect)$$

For example: let M be meningitis and S be stiff neck. Also:

$\mathrm{P}(M) = 0.0001$

$\mathrm{P}(S) = 0.1$

$\mathrm{P}(S \mid M) = 0.8$

Then, $\mathrm{P}(M \mid S)$:

$$\begin{aligned} &= \mathrm{P}(S \mid M) * \mathrm{P}(M) \;/\; \mathrm{P}(S) \\ &= 0.8 * 0.0001 \;/\; 0.1 \\ &= 0.0008 \end{aligned}$$

Hence note, posterior probability of meningitis, given that you have a stiff neck, is still very small!