Bayesian Cognitive Science

Cognitive Science views the brain as an *Information Processor*:

- Information comes from the senses, language, memory etc.
- Information is typically uncertain / noisy.
- We need to reason about the past to help with the present and future.

Probability and Information Theory is a natural way to think about CogSci.

Playing Tennis

Suppose you are playing tennis:

- You know how quickly you can move.
- You have an idea how your partner will server.
- How can you anticipate the best action to take?

Playing Tennis

We can think of the world as being a *state*:

- A state encodes where the ball will bounce.

We can connect to the world using *sensory input*:

- Sensory input means watching the opponent.

Playing Tennis

Putting the two together:

$$P(\text{state} \mid \text{sensory input}) = \frac{P(\text{sensory input} \mid \text{state})P(\text{state})}{P(\text{sensory input})}$$

This allows us to combine together what we see with that we believe.

Experimental results suggest that people learn the prior and combine with sensory evidence is a similar manner.

Eating Puffer Fish

Puffer fish is a delicacy. But (WikiPedia):

> (contains) a powerful neurotoxin that can cause death in nearly 60% of the humans that ingest it. A human only has to ingest a few milligrams of the toxin for a fatal reaction to occur. Once consumed the toxin blocks the sodium channels in the nervous tissues, ultimately paralyzing the muscle tissue.

Eating Puffer Fish

People like eating Puffer Fish, yet have to consider the possibility of being poisoned.

- Can we reason about the cost of eating Puffer Fish against the yummy taste?
- We wish to select some action which has the lowest average cost over all possible states.
- Decision Theory allows us to reason about taking optimal actions.

Eating Puffer Fish

Decision theory connects actions with probabilities:

- $L(X,Y)$ is a *loss function*.
- A loss function characterises the cost of taking action X in state Y.
  - L(eat, poisoned): the cost of eating bad fish.
  - L(eat, safe): the cost of eating good fish.

Eating Puffer Fish

We need to consider all possible states:

$$E(\text{action}) = \sum_{\text{state}} L(\text{action}, \text{state}) P(\text{state} \mid \text{action})$$

- Suppose we believe that the cost of eating bad fish is 5000
- And believe that the cost of eating safe fish is 0.
- $P(\text{poisoned} \mid \text{eat}) = \frac{1}{10{,}000}$

Should we eat the fish?

Eating Puffer Fish

The expected loss of eating fish is then:

$$\begin{aligned} E(\text{eat}) = \quad & L(\text{eat, poisoned})P(\text{poisoned} \mid \text{eat}) + \\ & L(\text{eat, safe})P(\text{safe} \mid \text{eat}) \\ & = -0.4999 \end{aligned}$$

- If we do nothing, then the loss is zero.

We should therefore eat the fish.

Decision Theory

- Decision theory allows us to reason about the relationship between perceived costs and uncertainty.
- DT has been applied to neural processing.

Occam's Razor

One theory of human learning is that we try to find simple descriptions:

- The World rests on a tortoise, which swims in an ocean ...
- The World is a rock in space.

Occam's Razor:

> All things being equal, the simplest solution tends to be the best one.

How can we formalise simplicity?

Occam's Razor

Information Theory considers *compressing* items:

- If an item X occurs with probability $P(X)$, then an optimal code will use $l(x) = -\log P(X)$ units to represent it.
- $l(x)$ is the *description length* of x.
  - Suppose the letter e has $P(e) = 1/5$ and z has $P(z) = 1/100$.
  - $l(e)$ = 1.6 units, $l(z)$ = 4.6 units
- The complexity of a theory is then equivalent to the description length of that theory.

Occam's Razor

Highly likely theories will receive short description lengths.

- An empty theory will have a minimally short description!
- We also need to consider how well the theory accounts for the data (…*All things being equal*).
- The likelihood is a natural way to talk about the data in terms of a theory: $P(D \mid M)$.
- $l(P(D \mid M))$ gives us the length of the data encoded in the model.

High likelihoods compactly describe the data.

Occam's Razor

Putting it together:

- Select a compact model, which describes the data simply:

$$L(M) + L(D \mid M)$$

- This shows the connection between Bayes Theorem and simplicity.
- Much CogSci can be seen in terms of simplicity:

Process	Data	Code
Language learning	Linguistic input	Grammars
Low-level preception	Sensory input	Filters in early vision
Ants	Paths to food	tactile contact between ants

Summary

Probability is a rich language for CogSci:

- Bayes allows us to talk about combining together existing knowledge with our current state-of-affairs.
- Decision theory allows us to reason about subjective costs and uncertainty.
- Information Theory lets us talk about simplicity in a formal manner.

Final comment: do we actually think using probabilities, or is it just a metaphor?