A Little Opening Fun

Three prisoners, A, B, and C, are locked in their cells. Everyone knows that one of them will be executed the next day and that the other two will be pardoned. Only the governor knows which one will be executed, and that is how he likes it.

Prisoner A asks the prison guard a favor: “Please ask the governor who will be executed, and then take a message to one of my friends, either B or C, to let him know that he will be pardoned in the morning.” The guard agrees to do just that.

Later, the guard stops by A’s cell. He tells A that he gave the pardon message to prisoner B.

What are A’s chances of being executed,
given this information?

Try to answer this <u>logically</u> or <u>mathematically</u>,
not by energetic waving of hands.

Acting—and Thinking—
in the Face of Uncertainty

How are we forced to act from a position of uncertainty?

- inadequate sensors and effectors
- the nature of the environment
- the pragmatics of knowledge representation

The consequences of these practicalities:

- Almost certainly, data are uncertain.
- Conclusions are almost certainly uncertain.
- Actions taken are possibly inappropriate.
- The results of actions are almost certainly uncertain.
- The results of learning are uncertain.

Some side effects of these unfortunate realities:

- Logic and rules seem like burdensome knowledge representations.

- Intelligent agents may want to adopt a more "economic" view of the world, basing its decisions on *preferences* and *utilities* rather than on inferences.

Probability as a Tool

How can we augment a representation based in logic to reflect this broader view?

- Assertions are contingent.
- Each statement has an associated probability that denotes the agent's *degree of belief* in the assertion.

So, instead of saying:

The patient has a cavity.

we say:

There is an 80% chance that the patient has a cavity.

Probability as a Tool

An agent's degree of belief in an assertion can change over time:

- before seeing any evidence
  prior probability

$$P(x)$$

- after seeing evidence **e**
  posterior probability

$$P(x|e)$$

.

Imagine the case of the dental patient:

- before seeing any evidence
  $P(x) = 0.2$
- after evidence: phone call on tootache
  $P(x|e_1) = 0.7$
- after evidence: visual exam
  $P(x|e_1 \wedge e_2) = 0.95$
- after evidence: X-ray
  $P(x|e_1 \wedge e_2 \wedge e_3) = 0.999$

An Exercise

I have to fly to Phoenix on February 25. My internal planner has created a plan that is *99% certain* to get me to the airport on time for my flight.

Should I use this plan or try to find another one?

- If yes, why?
- If no, why not?

Probabilities are meaningless out of context.

- What is the cost of executing the plan?
- How averse to risk am I?
- How easily can I generate alternative plans?

This is an ***economic decision***.

Probability and Rationality

Probabilities are part of a bigger picture.

In the real world, agents have preferences for some states of the world over other states of the world.

Consider some examples:

- Some people donate 10% of their gross incomes to charity.
- You are playing in the last round of a chess tournament. You have 4.5 points, and your next closest competitor has 4.0 points. The winner of the tournament earns $10,000, and second place wins $5,000.
- You are playing Garry Kasparov a game of chess. It offers a draw in a position that you think you might be able to win.
- You are an injured Roger Clemens. You feel good enough to pitch in Game 5 today, though the team doctor tells you that there is a 50% chance that you will worsen your injury by pitching today.

Probability and Rationality

Given that all reasoning is done in the context of uncertainty, **risk aversion** becomes an important preference.

Economists and AI folks use a notion of ***utility theory*** to measure the expected outcomes of an action and choose among them.

“acting rationally”
==
choosing the action with the highest expected value

Probability is merely a way to compute expectations.

Probability

`x`	an assertion
`P(x)`	degree of belief in `x`
`P(x | E)`	degree of belief in `x` given that we know `E`

Some of the relevant laws of probability:

- `0 <= P(x) <= 1 for all x`
- `P(true) == 1 and P(false) == 0`
- `P(a or b) == P(a) + P(b) - P(a and b)`
  `if a and b are independent`
- `P(x | E) == P(x and E) / P(E)`

Bayes' Law: Probability in Practice

We know that:

```
P(x | E) == P(x and E) / P(E)
```

So:

```
P(x and E) == P(x | E) * P(E)
```

But evidence is relative, so:

```
P(E and x) == P(E | x) * P(x)
```

And so:

```
P(E | x) * P(x) == P(x | E) * P(E)
```

And:

```
                P(E | x) * P(x)
P(x | E) == ----------------
                     P(E)
```

This is Bayes' Law, the work horse of most AI that uses probabilistic reasoning.

Bayes' Law at Work

You recently read that doctors have reported several cases of meningitis in your town. You wake up one morning with a sore throat, so you decide to call your doctor. He encourages you to come in for an appointment but calms your fears of having meningitis. It turns out that meningitis strikes only 1 in 50,000 people, and the probability that a person will have a sore throat if they do have meningitis is only 50%. And, besides, the chance that you will wake up with a sore throat on any given day is 5%.

What are the chances that you have the disease?

```
P( HaveSoreThroat | HaveDisease ) = 0.50
P( HaveDisease )                  = 0.00002
P( HaveSoreThroat )               = 0.05

P( HaveDisease | HaveSoreThroat ) =
```

$$\frac{P(HaveSoreThroat|HaveDisease) * P(HaveDisease)}{P(HaveSoreThroat)}$$

```
      = ( 0.50 * 0.00002 ) / 0.05
      = 0.0002
```

An Exercise

After your yearly check-up, the doctor has good news and bad news. The bad news is that you have tested positive for a serious disease, and that the test is 99% accurate. The good news that this is a rare disease, striking only 1 in 10,000 people.

Why is the good news good news?
What are the chances you have the disease?

```
P(     PositiveTest |     HaveDisease ) = 0.99
P( not PositiveTest | not HaveDisease ) = 0.99
P( HaveDisease )                        = 0.0001
P( PositiveTest )                       = ??????

P( HaveDisease | PositiveTest )         = ??????

P( PositiveTest )
  = P(PositiveTest|    HaveDisease) * P(    HaveDisease) +
    P(PositiveTest|not HaveDisease) * P(not HaveDisease)

P( PositiveTest | HaveDisease ) * P( HaveDisease )
  = 0.99 * 0.0001
  = 0.000099

P( PositiveTest | not HaveDisease ) * P( not HaveDisease )
  = 0.009999

P( PositiveTest ) = 0.000099 + 0.009999
P( HaveDisease | PositiveTest )
  = 0.000099 / (0.000099 + 0.009999) = 0.009804
```

An Exercise

So, what does this tell us?

- humans as probabilistic reasoners
- difficulty of gathering the data...

.