Uncertainty

CPS 170

Ron Parr

Why do we need uncertainty?

- Many proposed logic as the “language” of AI
- Problem: General logical statements are almost always false
- Truthful and accurate statements about the world would seem to require an endless list of *qualifications*
- How do you start a car?
- Call this “The Qualification Problem”

The Qualification Problem

- Is this a real concern?
- YES!
- Systems that try to avoid dealing with uncertainty tend to be brittle.
- Plans fail
- Plan failures result in logical contradictions
- Logical contradictions can cause huge problems

Probabilities

- Natural way to represent uncertainty
- People have intuitive notions about probabilities
- Many of these are wrong or inconsistent
- Most people don’t get what probabilities mean
- Finer details of this question still debated

Probability Solves the Qualification Problem

- Later: we define conditional probability
- For example, P(disease|symptom1)
- This defines the probability of a disease given that we have observed only symptom1
- The conditioning bar indicates that the probability is defined with respect to a particular state of knowledge, not as an absolute thing

Understanding Probabilities

- Initially, probabilities are “relative frequencies”
- This works well for dice and coin flips
- For more complicated events, this is problematic
- What is probability Bush will win a second term?
  - This event only happens once
  - We can’t count frequencies
  - Still seems like a meaningful question
- In general, all events are unique
- “Reference Class” problem

Probabilities and Beliefs

- Suppose I have rolled a die and hidden the outcome
- What is P(Die = 3)?

- Note that this is a statement about a *belief*, not a statement about the world
- The world is in exactly one state and it is in that state with probability 1.
- Assigning truth values to probability statements is very tricky business
- Must reference speakers state of knowledge

Belief and Determinism

- Ask the same question before rolling the die
- Your probabilities are the same, but mine are different (they now match yours)
- The world has changed, my state of knowledge has changed, but yours has not
- Changing the world doesn't change our perception of probability; only changing our knowledge does
- There is a sense in which probability can necessarily be about our beliefs *only*: Cached random calls

Frequentism and Subjectivism

- Frequentists hold that probabilities must come from relative frequencies
- This is a purist viewpoint
- This is corrupted by the fact that relative frequencies are often unobtainable
- Often requires complicated and convoluted assumptions to come up with probabilities
- Subjectivists: probabilities are degrees of belief
  - Taints purity of probabilities
  - Often more practical

Why probabilities are good

- Subjectivists: probabilities are degrees of belief
- Are all degrees of belief probability?
  - AI has used many notions of belief:
    - Certainty Factors
    - Fuzzy Logic
- Can prove that a person who holds a system of beliefs inconsistent with probability theory can be tricked into accepted a sequence of bets that is guaranteed to lose

What are probabilities?

- Probabilities are defined over random variables
- Random variables are usually represented with capitals: X,Y,Z
- Random variables take on values from a finite domain d(X), d(Y), d(Z)
- We use lower case letters for values from domains
- X=x asserts that the random variable X has taken on value x
- P(x) is shorthand for P(X=x)

Domains

- In the simplest case, domains are boolean
- In general may include many different values
- Most general case: domains may be continuous
- This introduces some special complications

Kolmogorov's axioms of probability

- 0<=P(a)<=1
- P(true) = 1; P(false)=0
- P(a or b) = P(a) + P(b) – P(a and b)
- Subtract to correct for double counting

- This is sufficient to completely specify probability theory for discrete variables
- Continuous variables need density functions

Atomic Events

- When several variables are involved, it is useful to think about atomic events
- An atomic event is a complete assignment to variables in the domain
  (compare with interpretation in logic)
- Atomic events are mutually exclusive
- Exhaust space of all possible events
- For n binary variables, how many unique atomic events are there?

Joint Distributions

- A joint distribution is an assignment of probabilities to every possible atomic event
- We can define all other probabilities in terms of the joint probabilities:

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

$$P(a) = P(a \wedge b) + P(a \wedge \neg b)$$

Why Probabilities Are Messy

- Probabilities are not truth-functional
- To compute P(a and b) we need to consult the joint distribution
  - sum out all of the other variables from the distribution
  - It is not a function of P(a) and P(b)
  - It is not a function of P(a) and P(b)
  - It is not a function of P(a) and P(b)
- This fact led to many approximations methods such as certainty factors and fuzzy logic (Why?)
- Neat vs. Scruffy…

Conditional Probabilities

- Ordinary probabilities for random variables: unconditional or prior probabilities
- P(a|b) = P(a AND b)/P(b)
- This tells us the probability of a **given that we know *only* b**
- If we know c and d, we can't use P(a|b) directly
- Annoying, but necessary to solve the qualification problem

Conditioning

- Suppose we know P(ABCDE) ← Joint
- Observe B=b, update our beliefs:

$$P(acde \mid b) = \frac{P(abcde)}{P(b)} = \frac{P(abcde)}{\sum_{ACDE} P(AbCDE)}$$

Condition with Bayes's Rule

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

Let's Play Doctor

- Suppose P(test_positive|hepatitis) = 0.99
- Patient tests positive
- What is probability patient has hepatitis?

$$P(h \mid tp) = \frac{P(tp \mid h)P(h)}{P(tp)}$$
$$= \frac{0.99 * 0.00001}{0.001} = 0.0099$$

Why Bayes's Rule is Helpful

- P(Disease|Symptom)=P(S|D)P(D)/P(S)
- P(S|D) is usually easily measure or estimated
- P(D), P(S) also somewhat easier to determine
- This is an example of *causal* reasoning
- Causal reasoning decomposes nicely

Independence

- Joint probability distributions don't scale
- Independence makes our lives easier
- If A and B are independent then
  – P(A and B) = P(A)P(B)
  – P(A|B) = P(A)
- Why does this help?