Artificial Intelligence Programming

Probability

Chris Brooks

Department of Computer Science

University of San Francisco

19-2: Uncertainty

- In many interesting agent environments, **uncertainty** plays a central role.
- Actions may have nondeterministic effects.
  - Shooting an arrow at a target, retrieving a web page, moving
- Agents may not know the true state of the world.
  - Incomplete sensors, dynamic environment
- Relations between facts may not be deterministic.
  - Sometimes it rains when it's cloudy.
  - Sometimes I play tennis when it's humid.
- Rational agents will need to deal with uncertainty.

19-3: Logic and Uncertainty

- We've already seen how to use logic to deal with uncertainty.
  - $Studies(Bart) \vee WatchesTV(Bart)$
  - $Hungry(Homer) \Rightarrow Eats(Homer, HotDog) \vee Eats(Homer, Pie)$
  - $\exists x Hungry(x)$
- Unfortunately, the logical approach has some drawbacks.

19-4: Weaknesses with logic

- Qualifying all possible outcomes.
  - "If I leave now, I'll be on time, unless there's an earthquake, or I run out of gas, or there's an accident ..."
- We may not know all possible outcomes.
  - "If a patient has a toothache, she may have a cavity, or may have gum disease, or maybe something else we don't know about."
- We have no way to talk about the likelihood of events.
  - "It's possible that I'll get hit by lightning today."

19-5: Qualitative vs. Quantitative

- Logic gives us a **qualitative** approach to uncertainty.
  - We can say that one event is more common than another, or that something is a possibility.
  - Useful in cases where we don't have statistics, or we want to reason more abstractly.
- Probability allows us to reason **quantitatively**
  - We assign concrete values to the chance of an event occurring and derive new concrete values based on observations.

19-6: Uncertainty and Rationality

- Recall our definition of rationality:
  - A rational agent is one that acts to maximize its performance measure.
- How do we define this in an uncertain world?
- We will say that an agent has a **utility** for different outcomes, and that those outcomes have a ***probability*** of occurring.
- An agent can then consider each of the possible outcomes, their utility, and the probability of that outcome occurring, and choose the action that produces the highest ***expected*** (or average) utility.
- The theory of combining preferences over outcomes with the probability of an outcome's occurrence is called ***decision theory***.

19-7: Basic Probability

- A probability signifies a **belief** that a proposition is true.
  - P(BartStudied) = 0.01
  - P(Hungry(Homer)) = 0.99
- The proposition itself is true or false - we just don't know which.
- This is different than saying the sentence is partially true.
  - "Bart is short" - this is **sort of** true, since "short" is a vague term.
- An agent's **belief state** is a representation of the probability of the value of each proposition of interest.

19-8: Random Values

- A random variable is a variable or proposition whose value is unknown.
- It has a domain of values that it can take on.
- These variables can be:
  - Boolean (true, false) - Hungry(Homer), isRaining
  - Discrete - values taken from a countable domain.
    - Temperature: <hot, cool, mild>, Outlook: <sunny, overcast, rain>
  - Continuous - values can be drawn from an interval such as [0,1]
    - Velocity, time, position
- Most of our focus will be on the discrete case.

19-9: Atomic Events

- We can combine propositions using standard logical connectives and talk about conjunction and disjunction
  - $P(Hungry(Homer) \land \neg Study(Bart))$
  - $P(Brother(Lisa, Bart) \lor Sister(Lisa, Bart))$
- A sentence that specifies a possible value for every uncertain variable is called an **atomic event**.
  - Atomic events are mutually exclusive
  - The set of all atomic events is exhaustive
  - An atomic event predicts the truth or falsity of every proposition
- Atomic events will be useful in determining truth in cases with multiple uncertain variables.

19-10: Axioms of Probability

- All probabilities are between 0 and 1. $0 \leq P(a) \leq 1$
- Propositions that are necessarily true have probability 1.
- Propositions that are unsatisfiable have probability 0.
- The probability of $(A \lor B)$ is $P(A) + P(B) - P(A \land B)$

19-11: Prior Probability

- The *prior probability* of a proposition is its probability of taking on a value *in the absence of any other information.*
  - P(Rain) = 0.1, P(Overcast) = 0.4, P(Sunny) = 0.5
- We can also list the probabilities of combinations of variables
  - $P(Rain \land Humid) = 0.1, P(Rain \land \neg Humid) = 0.1, P(Overcast \land Humid) = 0.2, P(Overcast \land \neg Humid) = 0.2, P(Sunny \land Humid) = 0.15, P(Sunny \land \neg Humid) = 0.25$
- This is called a *joint probability distribution*
- For continuous variables, we can't enumerate values
- Instead, we use a parameterized function.
  - $P(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$ (Normal distribution)

19-12: Inference with Joint Probability Distributions

- The simplest way of doing probabilistic inference is to keep a table representing the joint probability distribution.
- Observe each of the independent variables, then look up the probability of the dependent variable.

	Hum = High Sky = Overcast	Hum = High Sky = Sunny	Hum = Normal Sky = Overcast	Hum = Normal Sky = Sunny
Rain	0.1	0.05	0.15	0.05
¬Rain	0.2	0.15	0.1	0.2

19-13: Inference with Joint Probability Distributions

- We can also use the joint probability distribution to determine the **marginal probability** of the dependent variable by summing all the ways the dependent variable can be true.
  - $P(Rain) = 0.1 + 0.05 + 0.15 + 0.05 = 0.35$
- What is the problem with using the joint probability distribution to do inference?

19-14: Inference with Joint Probability Distributions

- We can also use the joint probability distribution to determine the **marginal probability** of the dependent variable by summing all the ways the dependent variable can be true.
  - $P(Rain) = 0.1 + 0.05 + 0.15 + 0.05 = 0.35$
- What is the problem with using the joint probability distribution to do inference?
- A problem with n independent Boolean variables requires a table of size 2^n

19-15: Conditional Probability

- Once we begin to make observations about the value of certain variables, our belief in other variables changes.
  - Once we notice that it's cloudy, $P(Rain)$ goes up.
- this is called *conditional probability*
- Written as: $P(Rain|Cloudy)$
- $P(a|b) = \frac{P(a \wedge b)}{P(b)}$
- or $P(a \wedge b) = P(a|b)P(b)$
  - This is called the *product rule*.

19-16: Conditional Probability

- Example: $P(Cloudy) = 0.3$
- $P(Rain) = 0.2$
- $P(cloudy \wedge rain) = 0.15$
- $P(cloudy \wedge \neg Rain) = 0.1$
- $P(\neg cloudy \wedge Rain) = 0.1$
- $P(\neg Cloudy \wedge \neg Rain) = 0.65$
  - Initially, $P(Rain) = 0.2$. Once we see that it's cloudy, $P(Rain|Cloudy) = P\frac{(Rain \wedge Cloudy)}{P(Cloudy)} = \frac{0.15}{0.3} = 0.5$

19-17: Independence

- In some cases, we can simplify matters by noticing that one variable has no effect on another.
- For example, what if we add a fourth variable $DayOfWeek$ to our Rain calculation?
- Since the day of the week will not affect the probability of rain, we can assert $P(rain|Cloudy, Monday) = P(rain|cloudy, Tuesday) ... = P(rain|cloudy)$
- We say that $DayOfWeek$ and $Rain$ are independent.
- We can then split the larger joint probability distribution into separate subtables.
- Independence will help us divide the domain into separate pieces.

19-18: Bayes' Theorem

- Often, we want to know how a probability changes as a result o an observation.
- Recall the Product Rule:
  - $P(a \wedge b) = P(a|b)P(b)$
  - $P(a \wedge b) = P(b|a)P(a)$
- We can set these equal to each other
  - $P(a|b)P(b) = P(b|a)P(a)$
- And then divide by $P(a)$
  - $P(b|a) = \frac{P(a|b)P(b)}{P(a)}$
- This equality is known as Bayes' theorem (or rule or law).