

#### Why do we need uncertainty?

- Reason: Sh\*t happens
- · Actions don't have deterministic outcomes
- Can logic be the "language" of Al???
  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
- Finding shortest path to goal isn't that great if the path doesn't really get you to the goal

# 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

#### **Understanding Probabilities**

- · Initially, probabilities are "relative frequencies"
- · This works well for dice and coin flips
- · For more complicated events, this is problematic
- What is the probability that the democrats will control Congress in 2008?
  - 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)?
- 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

# 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

# The Middle Ground

- · No two events are ever identical, but
- No two events are ever totally unique either
- What is probability that Hillary Clinton will be elected President?
  - Women have run for VP before
  - People like Hillary Clinton have run before
  - Have background knowledge about the electorate

# Why probabilities are good

- · Subjectivists: probabilities are degrees of belief
- Are all degrees of belief probability?
  - Al 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 (Dutch book)

## 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: RV 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 states in search)
- · 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 by *marginalization*:

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

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

## Example

- P(cold ^ headache) = 0.4
- P(¬cold ∧ headache) = 0.2
- P(cold  $\land \neg$  headache) = 0.3
- P(¬ cold ∧ ¬ headache) = 0.1
- What are P(cold) and P(headache)?

## Independence

- If A and B are independent:
   P(A ∧ B) = P(A)P(B)
- P(cold A headache) = 0.4
- P(¬cold ∧ headache) = 0.2
- P(cold \lambda headache) = 0.3
- P(¬ cold ∧ ¬ headache) = 0.1
- Are cold and headache independent?

# Independence

- If A and B are mutually exclusive: P(A v B) = P(A)+P(B) (Why?)
- Examples of independent events:
  - Duke winning NCAA, Dem. winning white house
  - Two successive, fair coin flips
  - My car starting and my stereo working
  - etc.

## 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...

# The Scruffy Trap

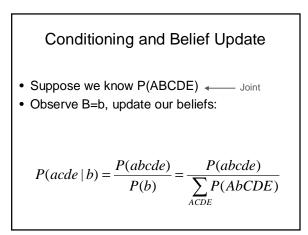
- Reasoning about probabilities correctly requires knowledge of the joint distribution
- This is exponentially large
- Very convenient to assume independence
- Assuming independence when there is not
- independence leads to incorrect answers • Examples:
  - ANDing symptoms
  - ORing symptoms

## **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 solves the qualification problem...

Probability Solves the Qualification Problem

- 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*



Condition with Bayes's Rule

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

# Example Revisited

- P(cold ∧ headache) = 0.4
- P(¬cold ∧ headache) = 0.2
- P(cold ^ headache) = 0.3
- $P(\neg \text{ cold } \land \neg \text{ headache}) = 0.1$
- What is P(cold|headache)?

#### Let's Play Doctor

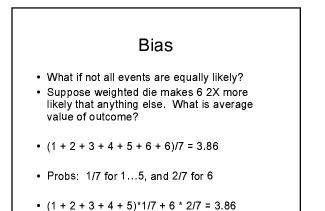
- Suppose P(cold) = 0.7, P(headache) = 0.6
- P(headache|cold) = 0.57
- What is P(cold|headache)?

$$P(c \mid h) = \frac{P(h \mid c)P(c)}{P(h)}$$
$$= \frac{0.57*0.7}{0.6} = 0.665$$

• IMPORTANT: Not always symmetric

#### Expectation

- Most of us use expectation in some form when we compute averages
- What is the average value of a die roll?
- (1+2+3+4+5+6)/6 = 3.5



## Expectation in General

- Suppose we have some RV X
- Suppose we have some function f(X)
- What is the expected value of f(X)?

$$E_{x} f(x) = \sum_{x} P(X) f(X)$$

