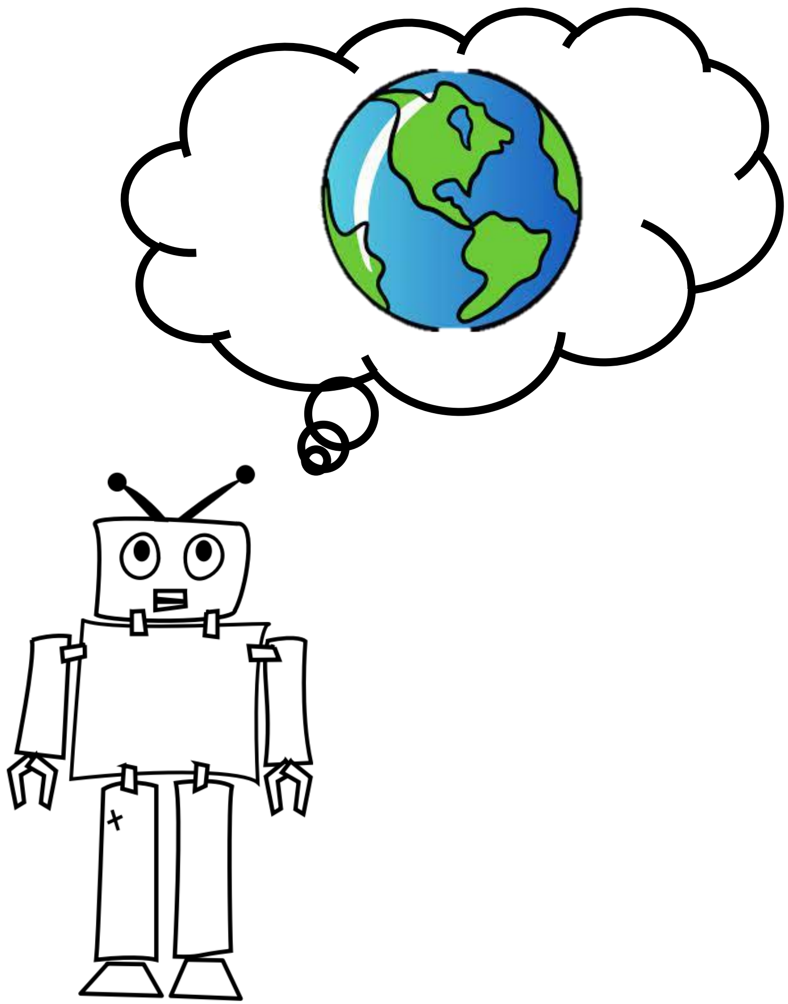
Artificial Intelligence

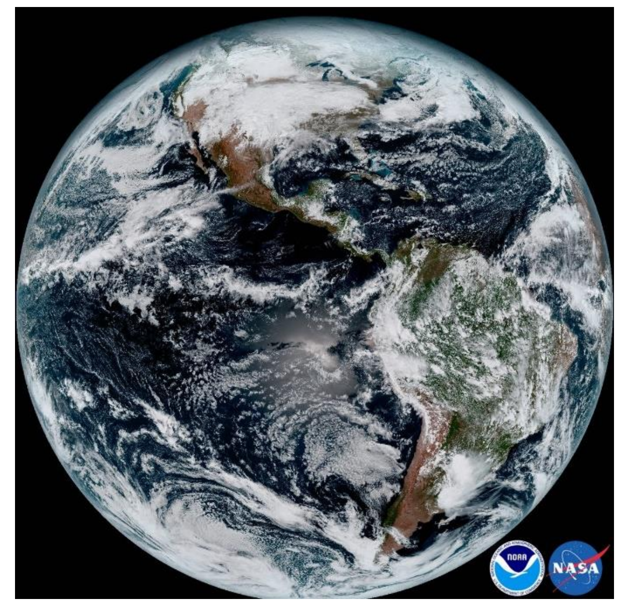
Steve James

Knowledge Representation & Reasoning

(Uncertain Knowledge)

Knowledge





Logic

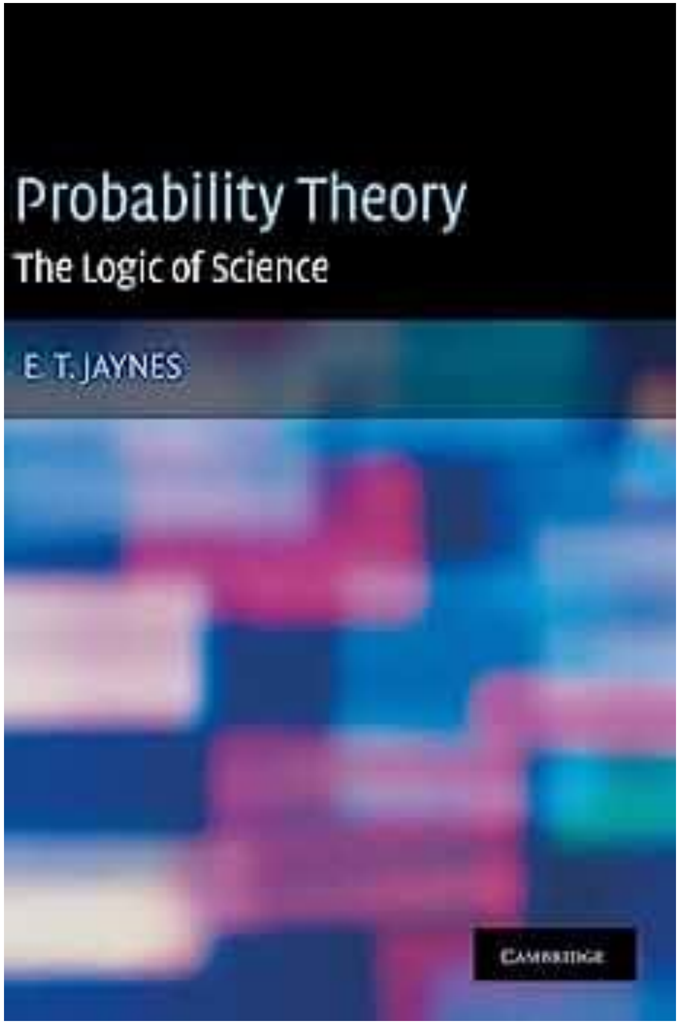
- Logical representations based on:
  - Facts about the world
  - Either true or false
  - We may not know which
  - Can be combined with logical connectives
- Logical inference based on what can be concluded with certainty

Logic is insufficient

- World is not deterministic
- No such thing as a fact
- Generalisation is hard

$$\forall x\ Fruit(x) \Rightarrow Tasty(x)$$

- Sensors/actuators are noisy
- Plans fail
- Models are imperfect
  - Learned models especially imperfect


Probability Theory
The Logic of Science
E. T. JAYNES
CAMBRIDGE

Probabilities

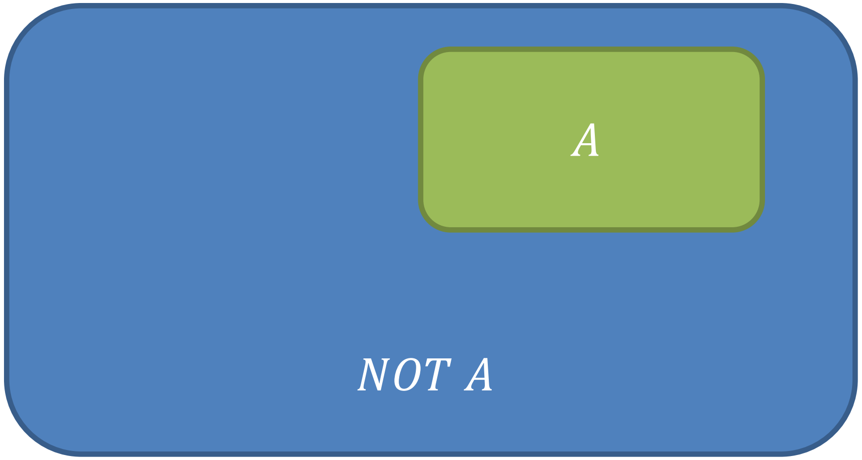
- Powerful tools for reasoning about uncertainty

- *Can prove that a person who holds a system of beliefs inconsistent with probability theory can be fooled*

- Not going to use them in the way you might expect

Relative frequencies

- Defined over events



- $P(A)$: probability random event falls in A rather that $Not\ A$
  - Works well for dice/coin flips

Relative frequencies

- But this feels limiting

- What is probability that South Africa wins the 2027 rugby world cup?
  - Meaningful question to ask
  - Can’t count frequencies (except naively)
  - Only really happens once
- In general, all events only happen once.

Probabilities and beliefs

- Suppose I flip a coin and hide outcome
  - What is $P(Heads)$?
- This is a statement about belief, not the world
  - World is in one state with probability 1
- Assigning truth values to probabilities is tricky
  - Must reference speaker's state of knowledge

- Frequentists: probabilities come from relative frequencies
- Subjectivists: probabilities are degrees of belief

Probabilities in AI

- No two events are identical, or completely unique
- Use probabilities as beliefs, but allow data (relative frequencies) to influence these beliefs

- In AI: probabilities reflect degrees of belief, given observed evidence.
- We use Bayes' Rule to combine prior beliefs with new data.

Example

- X: RV indicating winner of South Africa vs Australia game
- $d(X) = \{SA, Aus, Tie\}$
- A probability is associated with each event in the domain:
  - $P(X = SA) = 0.8$
  - $P(X = Aus) = 0.19$
  - $P(X = Tie) = 0.01$
- Note: probabilities over the entire event space must sum to 1

Joint probability distributions

- What to do when several variables are involved?

- Think about atomic events
  - Complete assignment of all variables
  - All possible events

- RVs: Raining, Cold (both boolean)

Proposition	**Value**
Cold	False
Raining	False

0.2

Proposition	**Value**
Cold	True
Raining	False

0.4

Proposition	**Value**
Cold	False
Raining	True

0.1

Proposition	**Value**
Cold	True
Raining	True

0.3

Raining	**Cold**	**Prob**
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

$$X \land Y$$

X	Y	P
True	True	1
True	False	0
False	True	0
False	False	0

$$X \lor Y$$

X	Y	P
True	True	0.33
True	False	0.33
False	True	0.33
False	False	0

$$\neg X$$

X	P
True	0
False	1

Joint probability distributions

- Probabilities to all possible atomic events (grows fast)

Raining	Cold	Prob
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

- Can define individual probabilities in terms of JPD:

$$P(Rain) = P(Rain, Cold) + P(Rain, not\ Cold) = 0.4$$

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

Joint probability distributions

- Simplistic probabilistic knowledge base:
  - Variables of interest $X_1, \ldots, X_n$
  - JPD over $X_1, \ldots, X_n$
  - Expresses all possible statistical information about relationships between the variables of interest
- Inference:
  - Queries over subsets of $X_1, \ldots, X_n$
  - $P(X_3)$
  - $P(X_3|X_1)$

Conditional probabilities

- What if you have a joint probability, and you acquire new data?

Raining	Cold	Prob
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

- My phone tells me it's cold
- What is prob of raining?

$$P(raining|cold)$$

Conditioning

$$P(X|Y)$$

- X is uncertain but Y is known (fixed, given)
- Ways to think about this:
  - X is belief, Y is evidence affecting belief
  - X is belief, Y is hypothetical
  - X is unobserved, Y is observed
- Soft version of implies:
  - $Y \Rightarrow X \approx P(X|Y) = 1$

Conditional probabilities

- We can write

$$P(a|b) = \frac{P(a\ and\ b)}{P(b)}$$

- This tells us the probability of a given only knowledge b
- This is a probability w.r.t a state of knowledge
  - $P(disease|symptom)$
  - $P(raining|cold)$
  - $P(SA\ wins|injury)$

Conditional probabilities

- $P(raining|cold) = \frac{P(raining\ and\ cold)}{P(cold)}$

- $P(cold) = 0.7$
- $P(raining\ and\ cold) = 0.3$

Raining	Cold	Prob
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

- $P(raining|cold) = 0.43$
- Note: $P(raining|cold) + P(not\ raining|cold) = 1$

Joint distributions are everything

- All you statistically need to know about $X_1, \dots, X_n$

- Classification
  - $P(X_1 | X_2, \dots, X_n)$

things you know

thing you want to know

- Co-occurrence
  - $P(X_a, X_b)$

how likely are these two things together?

- Rare event detection
  - $P(X_1, \dots, X_n)$

surprising event?

Joint probability distributions

- Grow very fast ☹
- Need to sum out other variables
- Might require lots of data
- Not a function of $P(A)$ and $P(B)$

Independence

- Critical property!
  - But rare
- If A and B are independent
  - $P(A \text{ } and \text{ } B) = P(A)P(B)$
  - $P(A \text{ } or \text{ } B) = P(A) + P(B) - P(A)P(B)$
- Independence: two events don’t affect each other
  - SA winning rugby world cup, Carlos Alcaraz winning Wimbledon
  - Two successive fair coin flips
  - It is raining and winning the lottery
  - Poker hand and date

Independence

- Are $Raining$ and $Cold$ independent?

Raining	Cold	Prob
True	True	0.3
True	False	0.1
False	True	0.4
False	False	0.2

- $P(Raining = True) = 0.4$
- $P(Cold = True) = 0.7$
- $P(Raining = True, Cold = True) =?$

Independence

- If independent, can break JPD into separate tables

Raining	Prob
True	0.6
False	0.4

Cold	Prob
True	0.75
False	0.25

Raining	Cold	Prob
True	True	0.45
True	False	0.15
False	True	0.3
False	False	0.1

Independence is critical

- Much of probabilistic knowledge representation and machine learning is concerned with identifying and leveraging independence and mutual exclusivity.

- Independence is also rare
  - Is there a weaker type of structure we might be able to exploit?

Conditional independence

- A and B are conditionally independent given C if:
  - $P(A|B,C) = P(A|C)$
  - $P(A,B|C) = P(A|C)P(B|C)$

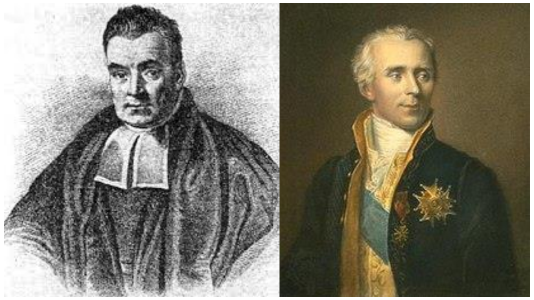
- If we know C, we can treat A and B as if they were independent
  - A and B might not be independent otherwise

Example

- Consider 3 random variables:
  - Temperature
  - Humidity
  - Season
- Temperature and humidity are not independent

- But they might be given the season
  - Season explains both and they become independent of each other

Bayes' rule



- Special piece of conditioning magic

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- If we have conditional $P(B|A)$ and we receive new data for B, we can compute new distribution for A (don't need joint)
  - As evidence comes in, revise belief.

Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' rule example

- Suppose
  - $P(disease) = 0.001$
  - $P(test|disease) = 0.99$
  - $P(test|no\ disease) = 0.05$

- What is $P(disease|test)$?

- $P(d|t) = \frac{P(t|d)P(d)}{P(t)} = \frac{0.99 \times 0.001}{P(t)} \approx 0.0194$

$$P(t) = P(t|d)P(d) + P(t|\neg d)P(\neg d)$$
$$= 0.99 \times 0.001 + 0.05 \times 0.999 = 0.05094$$

Bayes' rule example

- Suppose
  - $P(aliens) = 0.0001$
  - $P(digits\ of\ pi|aliens) = 0.95$
  - $P(digits\ of\ pi|not\ aliens) = 0.001$
- What is $P(aliens|digits\ of\ pi)$?

$$P(A|\pi) = \frac{P(\pi|A)P(A)}{P(\pi)}$$

$$\approx 0.087$$

$$P(A|\pi) = \frac{0.95 \times 0.0001}{P(\pi)}$$

$$P(\neg A|\pi) = \frac{P(\pi|\neg A)P(\neg A)}{P(\pi)}$$

$$\approx 0.913$$

$$P(\neg A|\pi) = \frac{0.001 \times 0.9999}{P(\pi)}$$

$$\frac{0.001 \times 0.9999}{P(\pi)} + \frac{0.95 \times 0.0001}{P(\pi)} = 1$$

$$P(\pi) = 0.0010949$$

Bayesian knowledge bases

- List of conditional and marginal probabilities
  - $P(X_1) = 0.7$
  - $P(X_2) = 0.6$
  - $P(X_3|X_2) = 0.57$

- Queries
  - $P(X_2|X_1)$?
  - $P(X_3)$?
- Less onerous than a JPD, but you may, or may not, be able to answer some questions

Probabilistic robotics