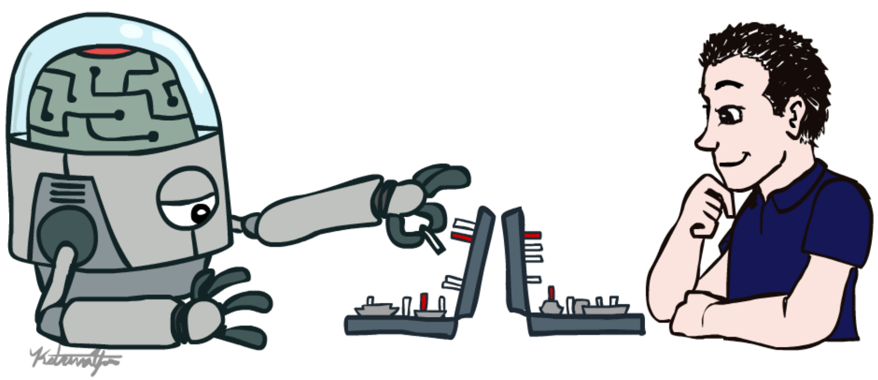
CSE 3521: Introduction to Artificial Intelligence



[Many slides are adapted from the UC Berkeley. CS188 Intro to AI at UC Berkeley and previous CSE 3521 course at OSU.]

THE OHIO STATE UNIVERSITY

What is Probability?

- “The probability the coin will land heads is 0.5”
  - Q: what does this mean?

What is Probability?

- "The probability the coin will land heads is 0.5"
  - Q: what does this mean?
- Interpretations:
  - Frequentist (Repeated trials)
    - If we flip the coin many times…

What is Probability?

- "The probability the coin will land heads is 0.5"
  - Q: what does this mean?
- Interpretations:
  - Frequentist (Repeated trials)
    - If we flip the coin many times…
  - Bayesian
    - We believe there is equal chance of heads/tails
    - Advantage: events that do not have long term frequencies

What is Probability?

- "The probability the coin will land heads is 0.5"
  - Q: what does this mean?
- Interpretations:
  - Frequentist (Repeated trials)
    - If we flip the coin many times…
  - Bayesian
    - We believe there is equal chance of heads/tails
    - Advantage: events that do not have long term frequencies

Q: What is the probability the polar ice caps will melt by 2050?

Probability Basic

- Begin with a set Ω
  - the sample space
    - e.g. 6 possible rolls of a die
    - $\omega \in \Omega$, is a sample point/possible world/atomic event
- A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$, s.t,
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{omega} P(\omega) = 1$
  - $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$
- An event A is a subset of Ω
  - $P(A) = \sum_{\{\omega \in A\}} P(\omega)$
  - $P(die\ roll < 4) = P(1) + P(2) + P(3) = P(A) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Random Variables

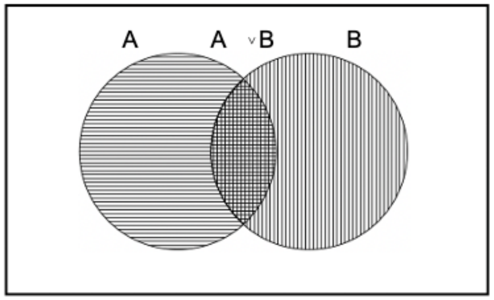
- A random variable is a function from sample points to some range e.g. the reals or Booleans
  - e.g. Odd(1) = True
- P induces a probability distribution for any random variable X
  - $P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$
  - e.g., $P(Odd = true) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Proposition

- Think of a proposition as the event (set of sample points) where the proposition is true
- Given Boolean random variables A and B
  - event a = set of sample points where $A(\omega)$ = true
  - event $\neg a$ = set of sample points where $A(\omega)$ = false
  - event $a \wedge b$ = points where $A(\omega)$ = true and $B(\omega)$ = true
- Often in AI applications, the sample points are defined by the values of a set of random variables,
  - i.e., the sample space is the Cartesian product of the ranges of the variables
- With Boolean variables, sample point = propositional logic model
  - e.g., A = true, B = f alse, or $a \wedge \neg b$.
- Proposition = disjunction of atomic events in which it is true
  - e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b) \Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Syntax

- The definitions imply that certain logically related events must have related probabilities
  - e.g., $P(a \lor b) = P(a) + P(b) - P(a \land b)$



- **de Finetti (1931):** An agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Why we should use probability

- Propositional or Boolean random variables
  - e.g., Cavity (do I have a cavity?)
  - Cavity = true is a proposition, also written cavity
- Discrete random variables (finite or infinite)
  - e.g., Weather is one of sunny, rain, cloudy, snow
    - Weather = rain is a proposition
  - Values must be exhaustive and mutually exclusive
- Continuous random variables (bounded or unbounded)
  - e.g., Temp = 21.6; also allow, e.g., Temp < 22.0.
- Arbitrary Boolean combinations of basic propositions

Prior probability

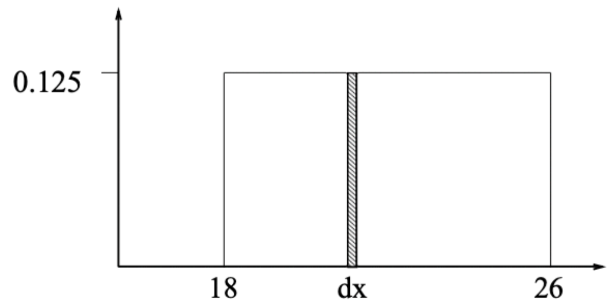
- Prior or unconditional probabilities of propositions
  - e.g., P (Cavity = true) = 0.1 and P (Weather = sunny) = 0.72
  - correspond to belief prior to arrival of any (new) evidence
- Probability distribution gives values for all possible assignments:
  - P(Weather) = 0.72, 0.1, 0.08, 0.1 (normalized, i.e., sums to 1)
- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables
  - i.e., every sample point
  - P(Weather, Cavity) = a 4 $\times$ 2 matrix of values

Weather	Sunny	Rain	Cloudy	Snow
Cavity= True	0.144	0.02	0.016	0.02
Cavity = False	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Probability for Continuous Varibales

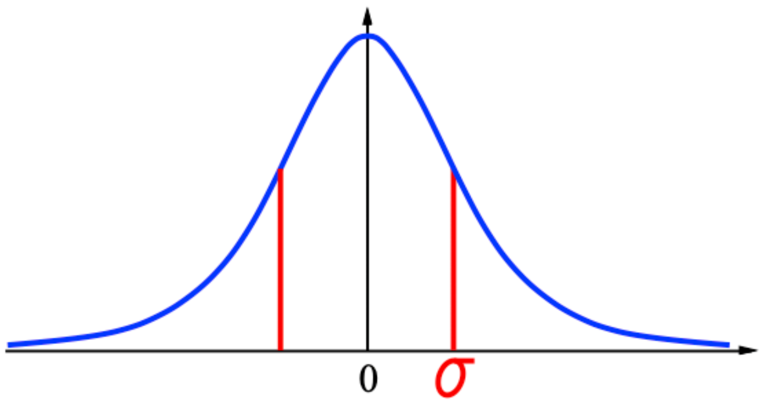
- Express distribution as a parameterized function of value:
  - P (X = x) = U [18, 26](x) = uniform density between 18 and 26



- Here P is a density; integrates to 1.
  - P (X = 20.5) = 0.125 really means
  - $_{dx}\lim \rightarrow 0$ P(20.5 ≤ X ≤ 20.5 + dx)/dx = 0.125

Gaussian Density

- $P(X) = \frac{1}{2\pi\sigma^2} e^{-\{x-u\}^2/2\sigma^2}$



Conditional Probability

- Conditional or posterior probabilities
  - e.g., P (cavity | toothache) = 0.8
  - i.e., given that toothache is all I know
    - NOT “if toothache then 80% chance of cavity”
- Notation for conditional distributions:
  - P(cavity | toothache) = 2-element vector of 2-element vectors)
  - If we know more, e.g., cavity is also given, then we have
  - P (cavity | toothache, cavity) = 1
- Note: the less specific belief remains valid after more evidence arrives
    - but it is not always useful
- New evidence may be irrelevant, allowing simplification
  - e.g.,P (cavity | toothache, 49℃ Weather) = P (cavity | toothache) = 0.8
  - This kind of inference, sanctioned by domain knowledge, is crucial

Inference by Enumeration

- Start with joint distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition, Φ, sum the atomic events where it is true
  - $P(\Phi) = \sum_{\omega:\omega \models \Phi} P(\omega)$

Inference by Enumeration

- Start with joint distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition, Φ, sum the atomic events where it is true
  - $P(\Phi) = \sum_{\omega:\omega \models \Phi} P(\omega)$
  - P (toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by Enumeration

- Start with joint distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition, Φ, sum the atomic events where it is true
  - $P(\Phi) = \sum_{\omega:\omega \models \Phi} P(\omega)$
  - P(cavity ∨ toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

Inference by Enumeration

- Start with joint distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition, Φ, sum the atomic events where it is true
  - $P(\Phi) = \sum_{\omega:\omega \models \Phi} P(\omega)$
  - P(no cavity | toothache) = $\frac{P(\neg cavity \wedge toothache)}{P(toothache)}$ = $\frac{0.016+0.064}{0.108+0.012+0.016+0.064}$ = 0.4

Normalization

- Start with joint distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Denominator can be viewed as a normalization constant α

P(cavity| toothache)
= α P(Cavity, toothache)
= α [P(Cavity, toothache, catch) + P(Cavity, toothache, ¬catch)]
= α [< 0.108, 0.016 > + < 0.012, 0.064 >]
= α < 0.12, 0.08 >
= < 0.6, 0.4 >

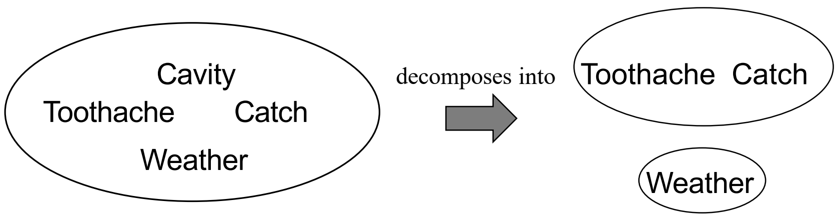
- General idea: compute distribution on query variable
  - by fixing evidence variables and summing over hidden variables

Inference by Enumeration

- Let X be all the variables. Typically, we want
  - the posterior joint distribution of the query variables Y given specific values e for the evidence variables E
- Let the hidden variables be $H = X - Y - E$
  - Then the required summation of joint entries is done by summing out the hidden variables:
  - $P(Y \mid E = e) = \alpha P(Y, E = e) = \alpha \Sigma_h P(Y, E = e, H = h)$
- The terms in the summation are joint entries because Y, E, and H together exhaust the set of random variables
- Obvious problems
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity
  2. Space complexity $O(d^n)$ to store the joint distribution
  3. How to find the numbers for $O(d^n)$ entries???

Independence

- A and B are independent iff
  - $P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)P(B)$



- P(toothache, catch, cavity, weather) = P(toothache, catch, cavity) P(weather)
- 32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$
  - Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent.
  - What to do?

Conditional Independence

- P(toothache, cavity, catch) has $2^3 - 1 = 7$ independent entries

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - $P(catch \mid toothache, cavity) = P(catch \mid cavity)$

- The same independence holds if I haven't got a cavity:
  - $P(catch \mid toothache, \neg cavity) = P(catch \mid \neg cavity)$

- Catch is conditionally independent of toothache given cavity:
  - $P(catch \mid toothache, cavity) = P(catch \mid cavity)$

- Equivalent statements:
  - $P(toothache \mid catch, cavity) = P(toothache \mid cavity)$
  - $P(toothache, catch \mid cavity) = P(toothache \mid cavity)\, P(catch \mid cavity)$

Conditional Independence

- Write out full joint distribution using chain rule:
  P(toothache, catch, cavity)
  = P(toothache| catch, cavity)P(catch, cavity)
  = P(toothache| catch, cavity)P(catch | cavity) P(cavity)
  = P(toothache| cavity)P(catch | cavity) P(cavity)

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

- Bayes' rule $P(a|b) = \frac{P(b|a)P(a)}{P(b)}$
- or in distribution form
- $P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \alpha P(X|Y)P(Y)$
- Useful for assessing diagnostic probability from causal probability:
  - $P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$
- e.g., let M be meningitis, S be stiff neck:
  - $P(M|S) = \frac{P(S|M)P(M)}{P(S)}$

Bayes' Rule Conditional Independence

- $P(cavity | toothache \wedge catch)$
  $= \alpha\, P(toothache \wedge catch | cavity) P(cavity)$
  $= \alpha\, P(toothache | cavity) P(catch | cavity) P(cavity)$

- This is an example of a naive Bayes model:
  - $P(cause, effect_1, \ldots, effect_n) = P(cause) \Pi_i P(effect_i | cause)$

- Total number of parameters is linear in n