Uncertainty

CS 3600
Intro to Artificial Intelligence

What have we done so far

Planning

Assumptions: Fully observable, fully known, fully deterministic environments

Solution: sequence of actions (plan) [Up, Up, Right, Right, Right]

Markov Decision Processes

Assumptions: Fully observable, fully known, stochastic actions environments

Solution: state to action map (policy) $\pi(s_1)$ = Right, $\pi(s_2)$ = Up, $\pi(s_3)$ = Up

Reinforcement Learning

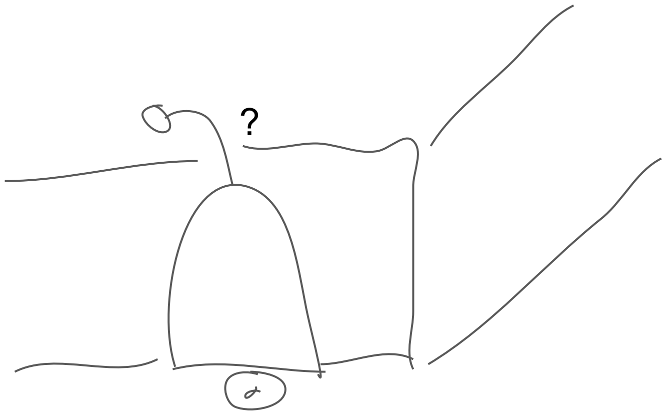
Assumptions: Fully observable, partially known, stochastic actions environments

Solution: state to action map (policy)

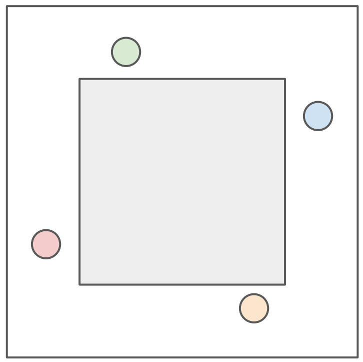
Sensor uncertainty

We’ve been assuming we can tell what state we’re currently in with perfect accuracy. This is often not true!

- Is it going to rain in the next hour?
- Do I have allergies or a cold?
- How much time do I have before my battery runs out?



Probabilities to the rescue!
S = {‘loc’:’A’, ‘A-clean’:True, ‘B-clean’:False} vs
S = {‘loc-is-A’:0.5, ‘A-clean’:0.9, ‘B-clean’:0.2}



We can ask questions like: “What’s the probability that both A and B are clean?”

Probabilistic Inference

Represent components of state as **Random Variables**:

$$p(X_i = \textsc{True}) = p, \quad 0 \leq p \leq 1$$

Random Variables can be **discrete** or continuous

$$\begin{aligned} \text{Raining} &\in \{\textsc{True}, \textsc{False}\} \\ \text{Battery} &\in [0, \infty) \end{aligned}$$

The function that determines the **probability** value is called the "distribution" of that RV

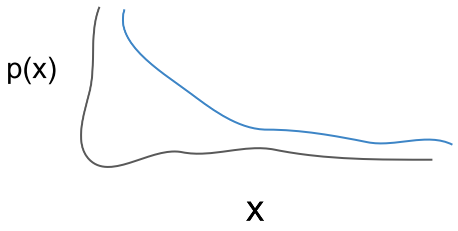
How do we know the distribution? Two ways:

Sampling by observing many times

- Expected utility from RL

Provided as part of the problem description

- p(Raining) = <0.1, 0.9>
- p(Battery > x) = e^{-x}



p(Raining=T)	p(Raining=F)
0.1	0.9

Rules of Probability

1. The probability that an RV takes on some value is always between 0 and 1

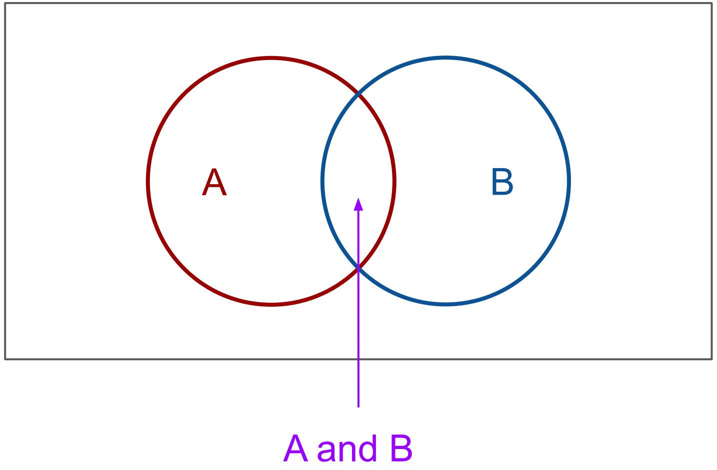
$$0 \leq p(X = x) \leq 1$$

2. Probability of deterministic events

$$p(\textsc{True}) = 1, \quad p(\textsc{False}) = 0$$

3. Additivity

$$p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B)$$



$$p(A = \textsc{True} \vee A = \textsc{False}) = p(A = \textsc{True}) + p(A = \textsc{False}) - p(A = \textsc{True} \wedge A = \textsc{False})$$
$$1 = p(A = \textsc{True}) + p(A = \textsc{False}) - 0$$
$$1 - p(A = \textsc{False}) = p(A = \textsc{True})$$

Interactions between Random Variables

With multiple interacting RVs, the probability distribution that includes **all** of them together is called the **joint distribution**

<u>Notation</u>
p(X): distribution, function or table
p(X=x): probability, single number

Example: p(Cavity, Toothache, Catch)

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Facts: table sums to 1, all combinations of all RVs, one cell per "configuration"

p(Toothache=F, Cavity=F, Catch=F) = 0.576

Events & Marginalization

- An **event** is a setting of some subset of random variables
- You can use the joint distribution to compute the probability of any event by **adding up** all the table entries that correspond with the given configuration

Marginalization

	Tooth		¬Tooth	
	Cat	¬Cat	Cat	¬Cat
Cav	0.108	0.012	0.072	0.008
¬Cav	0.016	0.064	0.144	0.576

p(Cav=T) = 0.108 + 0.012+0.072+0.008
= 0.2

p(Cav=T **or** Tooth=T) = 0.108+0.012+0.072+0.008 +0.16 + 0.064 = 0.28

p(Cav=T **and** Tooth=T) = 0.108 + 0.012 = 0.12

Conditional Probability (1)

- When one Random Variable **impacts** the value of another
  "If X=v, what is the probability of Y=w?"

$$p(Y = w \mid X = v) = \frac{p(Y = w \text{ and } X = v)}{p(X = v)}$$

"Conditioned variable"

<u>Notation</u>
p(Y|X=v): distribution, function or table
p(Y=w|X=v): probability, single number

Conditional Probability (2)

Typically, an agent wants to know conditional probabilities

$$p(\underbrace{S = s}_{\text{State}} \mid \underbrace{E = e}_{\text{Evidence}})$$

	Tooth		¬Tooth	
	Cat	**¬Cat**	**Cat**	**¬Cat**
Cav	0.108	0.012	0.072	0.008
¬Cav	0.016	0.064	0.144	0.576

Example:

p(Cav=T | Tooth=T) = p(Cav=T **and** Tooth=T)/p(Tooth=T)

= (0.108+0.012)/(0.108+0.012+0.016+0.064) = 0.6

p(Cav=F|Tooth=T) = (0.016+0.064)/(0.108+0.012+0.016+0.064) = 0.4

Normalization

p(Cav=T | Tooth=T) = 0.6
p(Cav=F | Tooth=T) = 0.4

Notice that the two probabilities summed to 1 across Cavity (**not** Toothache)

In general, we don't have to know/compute the denominator, we can "normalize"

$$p(X \mid E=e) = \frac{p(X, E=e)}{p(E=e)} = \alpha \cdot p(X, E=e)$$

$$1 = \alpha \cdot p(X = \textsc{True}, E=e) + \alpha \cdot p(X = \textsc{False}, E=e)$$

$$1 = \alpha \left[p(X = \textsc{True}, E=e) + p(X = \textsc{False}, E=e)\right]$$

$$\alpha = \frac{1}{\left[p(X = \textsc{True}, E=e) + p(X = \textsc{False}, E=e)\right]}$$

Hidden Variables

Frequently our agent won't have settings for **all** of the random variables.

Solution? Sum them out!

$$p(X \mid E = e) = \alpha \cdot p(X, E = e) = \alpha \sum_{h \in H} p(X, E = e, H = h)$$

In the previous example, Catch was a **hidden** variable:

p(Cav=T|Tooth=T) = αp(Cav=T,Tooth=T,Cat=T) + αp(Cav=T,Tooth=T,Cat=F)

Example - marginalization and normalization

What's the probability you have a cavity if you don't have a toothache?

p(Cav|Tooth=F) = α p(Cav,Tooth=F) = α $\sum_{h=\{T,F\}}$ p(Cav,Tooth=F,Catch=h)
= α[p(Cav, Tooth=F, Cat=T) + p(Cav, Tooth=F, Cat=F)]
= α[<0.072, 0.144> + <0.008, 0.576>]
= α <0.08, 0.72>

Must sum to 1, so
α(0.08+0.72) = α(0.8) = 1
α = 1.25

p(Cav|Tooth=F) = <0.1, 0.9>

	Tooth		¬Tooth	
	Cat	¬Cat	Cat	¬Cat
Cav	0.108	0.012	0.072	0.008
¬Cav	0.016	0.064	0.144	0.576

Conditioning

- Given full joint probability, agent can compute anything about the RVs!
- Usually, agent only has access to some conditional probabilities, not their joint distribution
- Can get around this by marginalizing and leveraging the definition of conditional probability

Ex: How can we get p(X) (marginal) if we don't know p(X,Y), just p(X|Y) and p(Y)?

Cond. Prob. def.

$$p(X \mid Y) = \frac{p(X,Y)}{p(Y)}$$

$$p(X \mid Y) \cdot p(Y) = p(X,Y)$$

Marginalize Y

$$\sum_y p(X \mid Y = y) p(Y = y) = \sum_y p(X, Y = y) = p(X)$$

Independence

Two random variables are **independent** if and only if their joint probability is **the same** as the product of their marginals

$$A \perp B \iff$$

$$p(A, B) = p(A)p(B)$$

$$p(A \mid B) = p(A)$$

Similarly, two random variables can be **conditionally independent** given a third

$$A \perp B \mid C \iff$$

$$p(A, B \mid C) = p(A \mid C)p(B \mid C)$$

$$p(A \mid B, C) = p(A \mid C)$$

Independence Example

	Tooth		¬Tooth	
	Cat	¬Cat	Cat	¬Cat
Cav	0.108	0.012	0.072	0.008
¬Cav	0.016	0.064	0.144	0.576

Are Catch and Toothache independent?

p(Catch=T, Toothache=T) = 0.124

p(Cat=T)*p(Tooth=T)=(0.108+0.016+0.072+0.144)*(0.108+0.012+0.016+0.064)

=(0.34)*(0.2) = 0.068 != 0.124 (**NOT INDEPENDENT**)

What about conditioned on Cavity?

p(Cat=T,Tooth=T|Cav=T) = 0.54

p(Cat=T | Cav=T)*p(Tooth=T|Cav=T) = 0.9*0.6=0.54

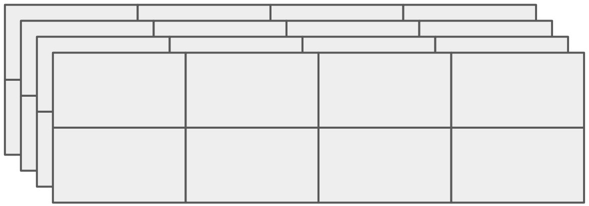
… also checks out for each setting of Cat, Cav, and Tooth

(**CONDITIONALLY INDEPENDENT given Cavity**)

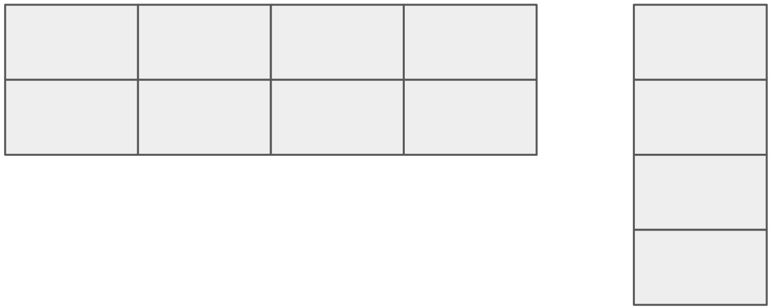
Factoring the joint probability with independence

- Suppose we add Weather as a RV to the Cav,Cat,Tooth model, where whether can take on 4 values: {Sunny, Rainy, Foggy, Snowy}

p(Cav,Tooth,Catch,Weather) = p(Cav,Tooth,Catch)*p(Weather)



2x2x2x4=32 cells



2x2x2 + 4=12 cells

Bayes Rule

Additional TRUE FACT about conditional probabilities

$$p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)}$$

Follows by plugging in definition of conditional probability

$$p(B \mid A) \cdot p(A) = p(A, B) = p(A \mid B)p(B)$$

This generalizes to more than two RVs, and gives us the **product rule**

$$p(A, B, C) = p(A, B \mid C) \cdot p(C) = p(A \mid B, C) \cdot p(B \mid C) \cdot p(C)$$

$$p(X_1, X_2, \ldots, X_d) = p(X_1 \mid X_2, \ldots, X_d) \cdot p(X_2 \mid X_3, \ldots, X_d) \cdot \ldots \cdot p(X_{d-1} \mid X_d) \cdot p(X_d)$$

Note: this is true for **any** ordering of the X_j!

Using Bayes Rule

Easy to measure

We can use Bayes Rule to “flip” the conditioning variable

$$p(\textsc{Cause} \mid \textsc{Effect}) = p(\textsc{Effect} \mid \textsc{Cause}) \frac{p(\textsc{Cause})}{p(\textsc{Effect})}$$

Difficult to measure

Easy to measure

No need to measure (normalize!)

Example: Medical Diagnosis “What’s the probability I have the flu given I have a cough?”

p(Flu|Cough=T) = α <p(Cough=T|Flu=T) p(Flu=T), p(Cough=T|Flu=F) p(Flu=F)>

Incorporating multiple pieces of evidence

"What's the probability of Cavity given Catch and Toothache?"

Applying bayes rule and normalizing:
p(Cav | Cat, Tooth) = α p(Tooth, Cat | Cav) p(Cav)

But since Catch and Tooth are conditionally independent given Cavity
p(Cav | Cat, Tooth) = α p(Tooth | Cav) p(Cat | Cav) p(Cav)

IF our **evidence** variables are conditionally independent of one another **given** the **cause** variable, we can significantly simplify things!

Summary and preview

Wrapping up

- To handle uncertain sensors, state now describes **probabilities**
- The joint probability distribution contains **all** the information we need to answer any question about any subset of the random variables it describes
- Marginalization, Normalization, and Bayes Rule are all “probability tricks” we can use to simplify/compute specific probabilities

Next time

- Detailed example: Avoid the wumpus!