Probability

CSE 473: Introduction to Artificial Intelligence

James Weichert (jpw@cs)
cs.uw.edu/473
acknowledgements

Welcome!
What is something you're uncertain about?
sli.do **#cse473**

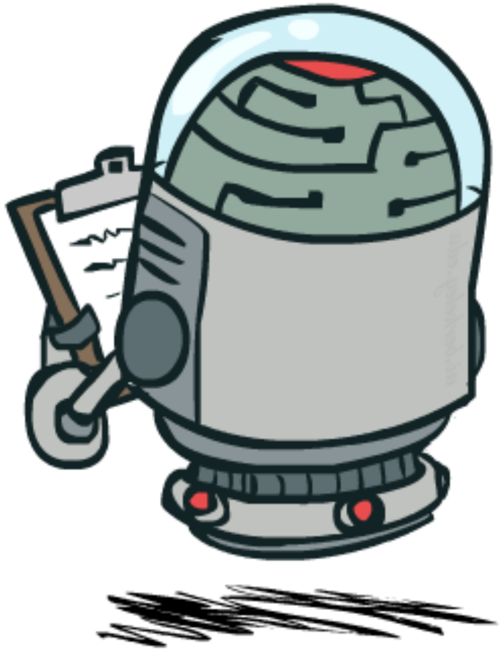
Administrivia

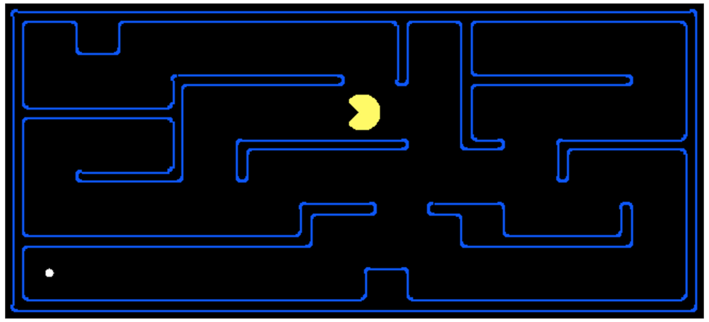
ANNOUNCEMENTS

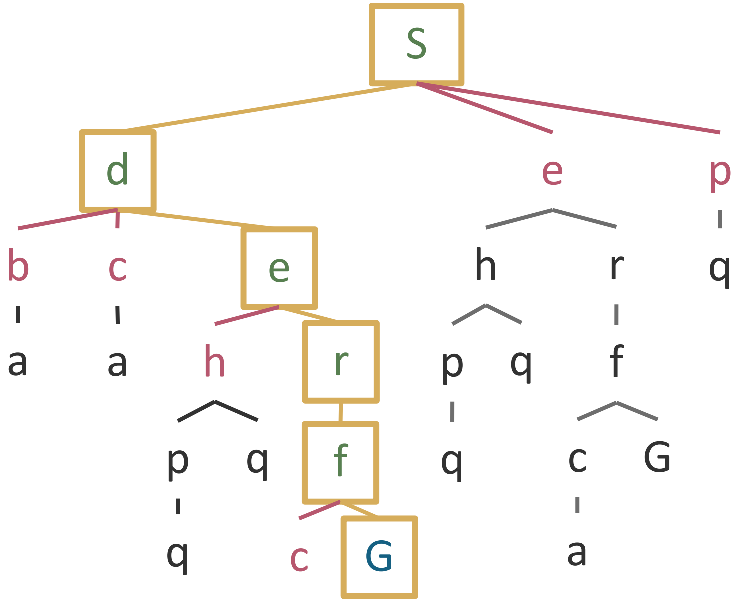
- **Test 1** scores released, regrade requests due Thursday

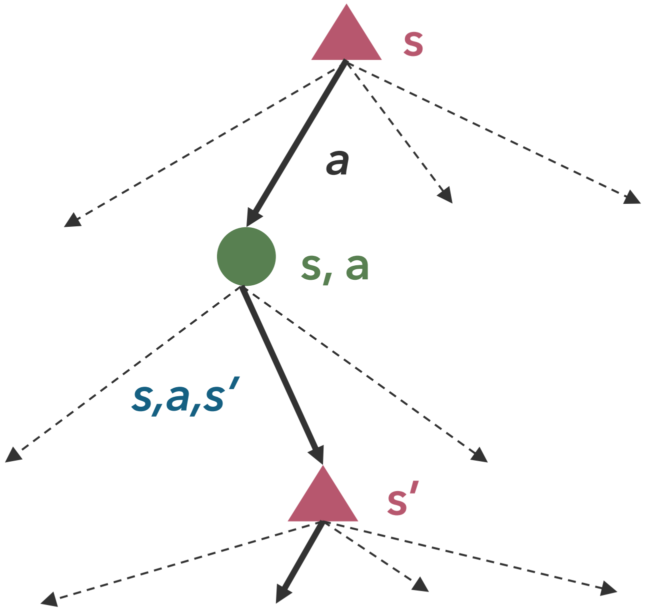
ASSIGNMENTS

- **Homework 2 due** Thursday (7.23)
- **Project 3 due** next Thursday (7.30)









Deterministic

Stochastic

Taking Stock

Randomness

CONTEXT

The real world is rife with uncertainty!

American League West

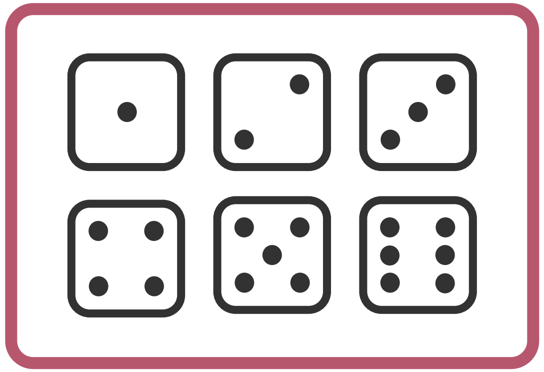
Team	W	L	Pct
Rangers	50	48	**.510**
Mariners	49	50	**.495**
Astros	47	53	**.470**
Athletics	42	56	**.429**
Angels	38	61	**.384**

Goal: *Maximize expected utility!*

Modeling Uncertainty

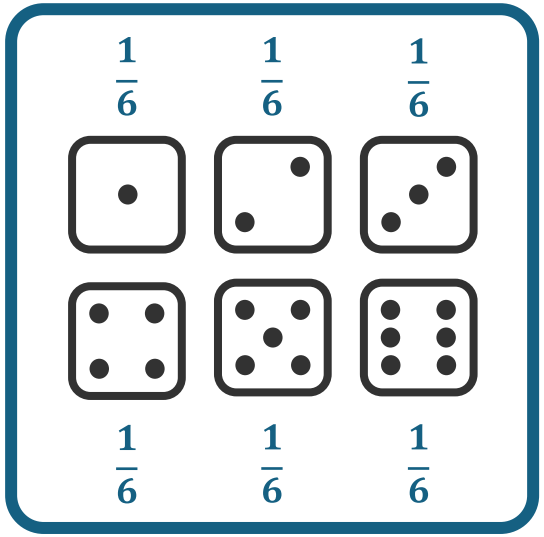
FOUNDATIONS

- Given set Ω of possible outcomes
- **Probability model** (distribution) assigns a number $P(\omega)$ for each outcome:



$$0 \leq P(\omega) \leq 1$$

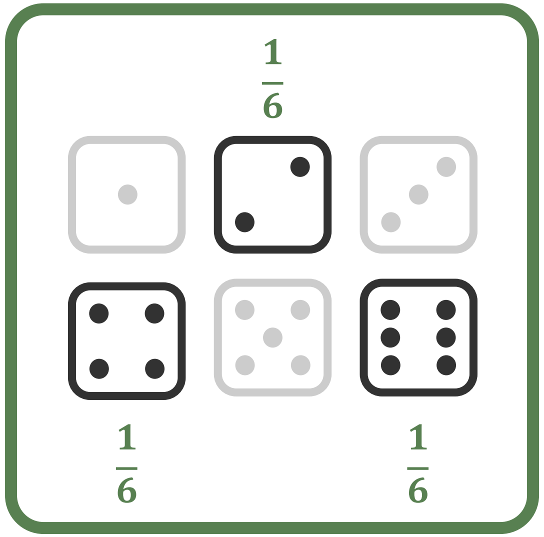
$$\sum_{\omega \in \Omega} P(\omega) = 1$$



Probabilities

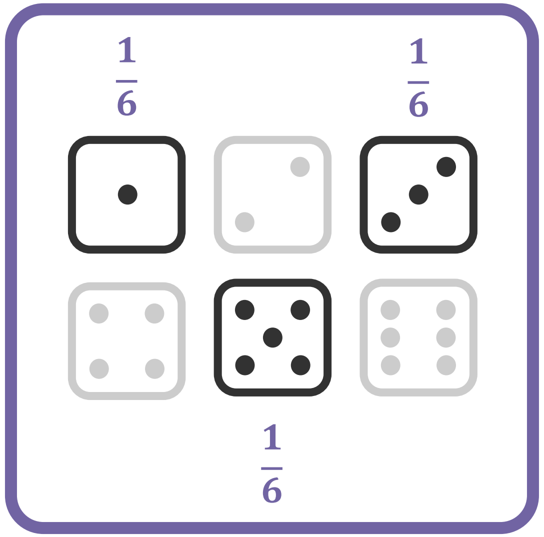
FOUNDATIONS

- An **event** is a subset of possible outcomes of **Ω**
- The probability of an event A is the sum of probabilities over the outcomes
  - All outcomes ω that satisfy the definition of event A



$P(\text{Even}) = \frac{1}{2}$

$$P(A) = \sum_{\omega \in A} P(\omega)$$



$P(\text{Odd}) = \frac{1}{2}$

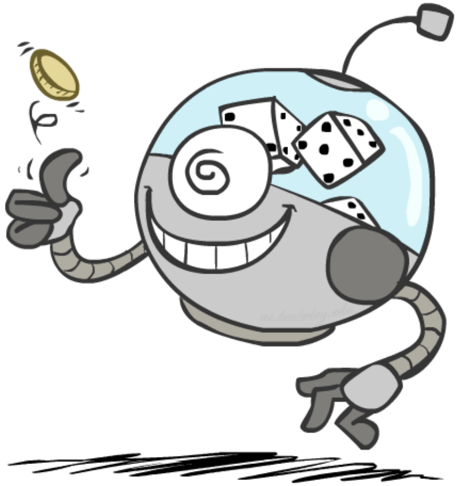
Random Variables

DEFINITION

A **random variable** represents an aspect of the world about which we are uncertain

- Range
  - Set of possible values $x \in X$
  - For example:
    - $\mathrm{Odd}(1) = \mathrm{true}, \mathrm{Odd}(6) = \mathrm{false}$
    - Range of Odd: $\{\mathrm{true}, \mathrm{false}\}$
- Probability Distribution $\boldsymbol{P}(X)$

$$\boldsymbol{P}(X = x) = \sum_{\{\omega:\, X(\omega)=x\}} P(\omega)$$



Questions?

live and on sli.do #cse473

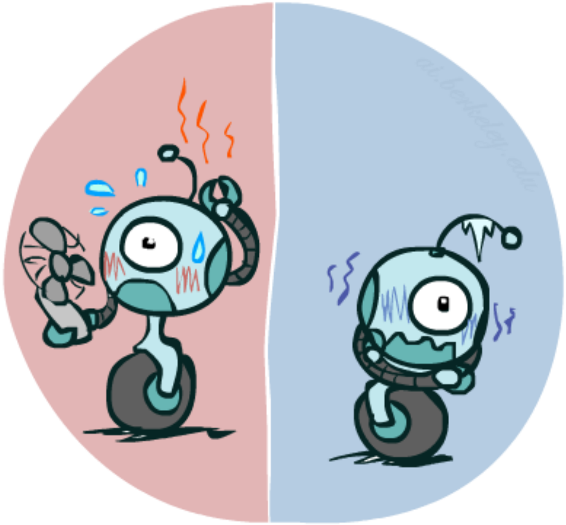
Probability Distributions

FOUNDATIONS

Probability distributions describe probabilities for *each* outcome in the range

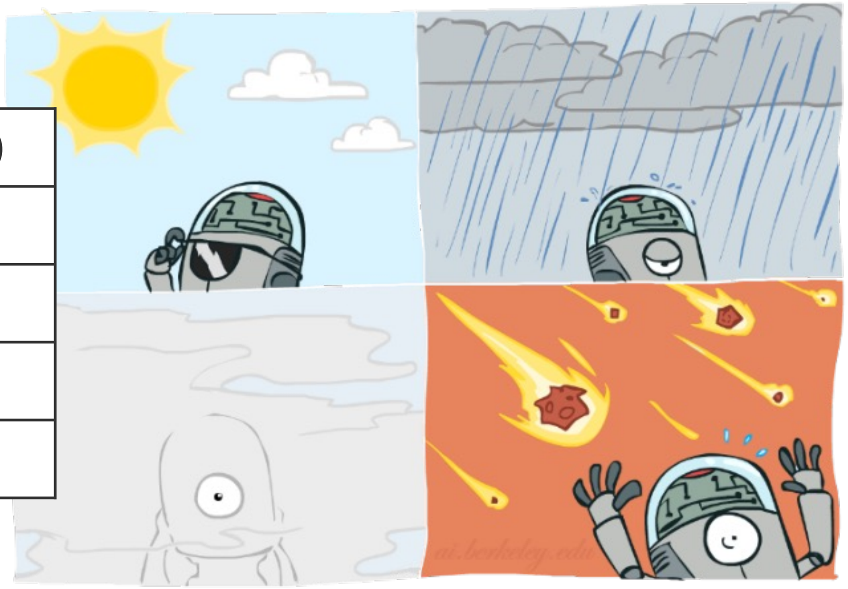
T: {hot, cold}

$T = t$	$P(t)$
hot	0.38
cold	0.62



W: {sun, rain, fog, meteor}

$W = w$	$P(w)$
sun	0.4
rain	0.5
fog	0.1
meteor	0.0



Joint Probabilities

FOUNDATIONS

- $P(A \cap B)$ is the probability of events A and B <u>both</u> occurring
- Joint distribution $P(X, Y)$
  - describes all combinations of values for the random variables X and Y

		Temperature	
		hot	cold
Weather	sun	0.3	0.1
	rain	0.05	0.45
	fog	0.03	0.07
	meteor	0.00	0.00

Size of distribution for n variables with range d:

$O(d^n)$

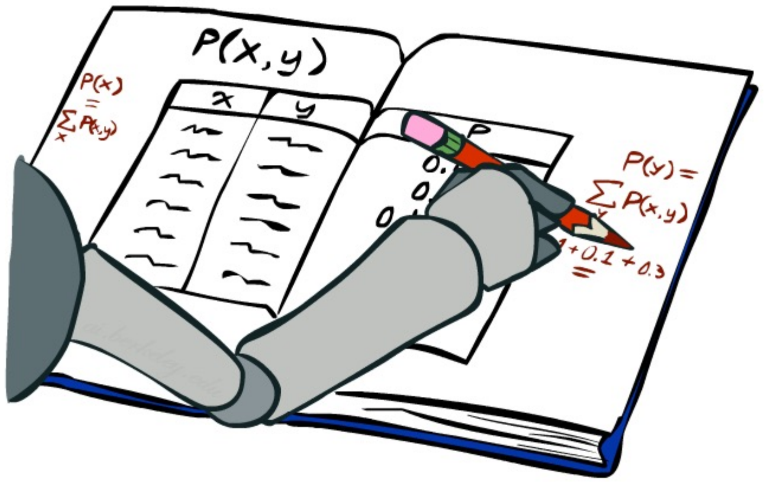
Marginal Distributions

FOUNDATIONS

Marginal distributions are *sub-tables* of joint distributions which eliminate variables

- A marginal probability can be calculated by summing over all values of other variables:

$$P(X = x) = \sum_{y \in Y} P(X = x, Y = y)$$



		Temperature		
		hot	cold	
Weather	sun	0.3	0.1	0.4
	rain	0.05	0.45	0.5
	fog	0.03	0.07	0.1
	meteor	0.00	0.00	0.0
		0.38	0.62	

$P(W)$

$P(T)$

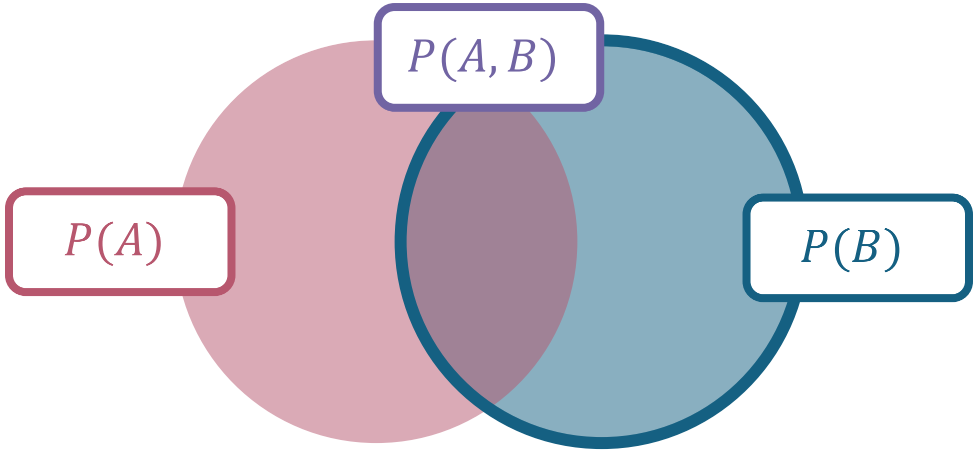
Conditional Probabilities

FOUNDATIONS

Conditional probabilities relate joint and marginal probabilities:

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

- The probability of A occurring ***given*** B occurs is the joint probability of <u>*both*</u> A and B occurring, divided (**normalized**) by the marginal probability of B occurring.



Determining Conditional Probabilities

EXAMPLE

Proportional, but not equal to, the joint probability with fixed evidence

$$\boldsymbol{P(W|T = \text{hot})}$$

0.79
0.13
0.08
0.00

$$= \frac{P(W, T = \text{hot})}{P(T = \text{hot})}$$

		Temperature	
		hot	cold
Weather	sun	0.3	0.1
	rain	0.05	0.45
	fog	0.03	0.07
	meteor	0.00	0.00

$$\boldsymbol{P(W|T = \text{cold})}$$

0.16
0.73
0.11
0.00

$$= \frac{P(W, T = \text{cold})}{P(T = \text{cold})}$$

Questions?

live and on sli.do #cse473

Probability Examples

EXAMPLE

answer on sli.do

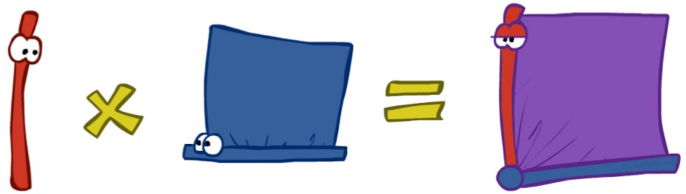
#cse473

X	Y	$P(X,Y)$
$+x$	$+y$	0.2
$+x$	$-y$	0.3
$-x$	$+y$	0.4
$-x$	$-y$	0.1

- $P(-x,-y)$?
- $P(+x)$?
- $P(-y)$?
- $P(-x \mid +y)$?

Using Conditional Probabilities

FOUNDATIONS



- The **Product Rule** calculates the joint probability:

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

  - If A and B are *independent*, then $P(A|B) = P(A) \Rightarrow P(A,B) = P(A) \cdot P(B)$

- The **Chain Rule** relates multiple conditional probabilities:

$$P(x_1, x_2, x_3) = P(x_3|x_2, x_1) \cdot P(x_2, x_2) = P(x_3|x_2, x_1) \cdot P(x_2|x_1) \cdot P(x_1)$$

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i|x_1, \dots, x_{i-1})$$

Probabilistic Inference

CONTEXT

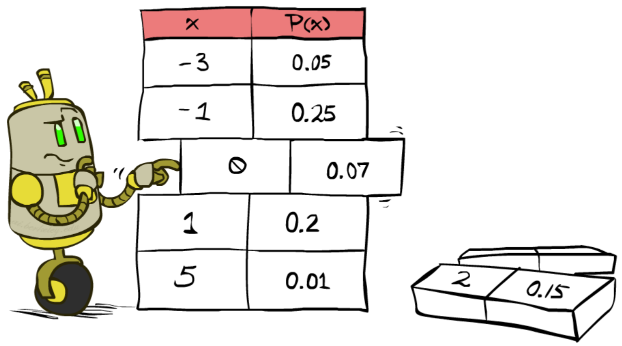
- We can *infer* things about the world by computing the relevant probability from a probability model
  - $\boldsymbol{P}(\boldsymbol{q}|\boldsymbol{e})$: the probability of **query q** given **evidence e**
  - These probabilities represent the agent's **beliefs** given the **evidence**
  - For example, $P(\text{airport on time} \mid \text{no accidents}) = 0.9$
- Beliefs may change with new evidence
  - $P(\text{airport on time} \mid \text{no accidents, 5am}) = 0.95$
  - $P(\text{airport on time} \mid \text{no accidents, 5am, raining}) = 0.8$
  - Observing new evidence causes beliefs to be updated

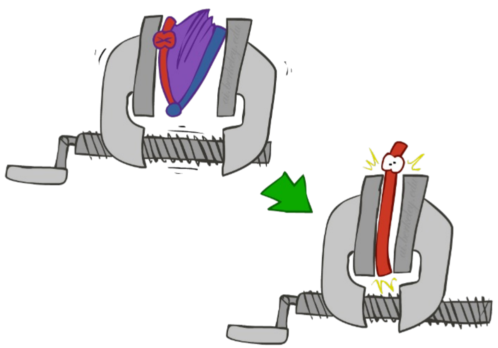
Inference by Enumeration

DEFINITION

- Calculate probability of query Q given:
  - Observations about evidence variables $E_1, \ldots, E_k = e_1, \ldots, e_k$
  - Hidden (unobserved) variables $H_1, \ldots, H_r$

Infer $P(Q|e_1, \ldots, e_k)$ from full probability model $P(X_1, \ldots, X_n)$





$$P(Q, h_1, \ldots, h_r, e_1, \ldots, e_k)$$

Select model entries consistent with evidence

$$P(Q, e_1, \ldots, e_k) = \sum_{h_1, \ldots, h_r} P(Q, h_1, \ldots, h_r, e_1, \ldots, e_k)$$

Sum out hidden variables to calculate joint probability

$$P(Q|e_1, \ldots, e_k) = \frac{P(Q, e_1, \ldots, e_k)}{P(e_1, \ldots, e_k)}$$

Normalize by joint probability of the evidence

Inferring the Weather

EXAMPLE

- $P(W)$?
  - $P(W) = \sum_{s \in S, t \in T} P(S = s, T = t, W)$
- $P(W|\text{summer})$?
  - $P(S = \text{summer}, W) = \sum_{t \in T} P(S = \text{summer}, T = t, W)$
  - $\alpha = \frac{1}{\sum_{w \in W} P(S=\text{summer}, W=w)}$
  - $P(W|\text{summer}) = \alpha \cdot P(S = \text{summer}, W)$
- $P(W|\text{summer}, \text{hot})$?
  - $P(S = \text{summer}, T = \text{hot}, W)$
  - $\alpha = \frac{1}{\sum_{w \in W} P(S=\text{summer}, T=\text{hot}, W=w)}$
  - $P(W|\text{summer}, \text{hot}) = \alpha \cdot P(S = \text{summer}, T = \text{hot}, W)$

S	T	W	$P(S,T,W)$
summer	hot	sun	0.35
summer	hot	rain	0.01
summer	hot	fog	0.01
summer	hot	meteor	0.00
summer	cold	sun	0.10
summer	cold	rain	0.05
summer	cold	fog	0.09
summer	cold	meteor	0.00
winter	hot	sun	0.10
winter	hot	rain	0.01
winter	hot	fog	0.02
winter	hot	meteor	0.00
winter	cold	sun	0.15
winter	cold	rain	0.20
winter	cold	fog	0.18
winter	cold	meteor	0.00

Problems with Inference by Enumeration

CONTEXT

Inference by Enumeration requires *a lot* of enumeration…

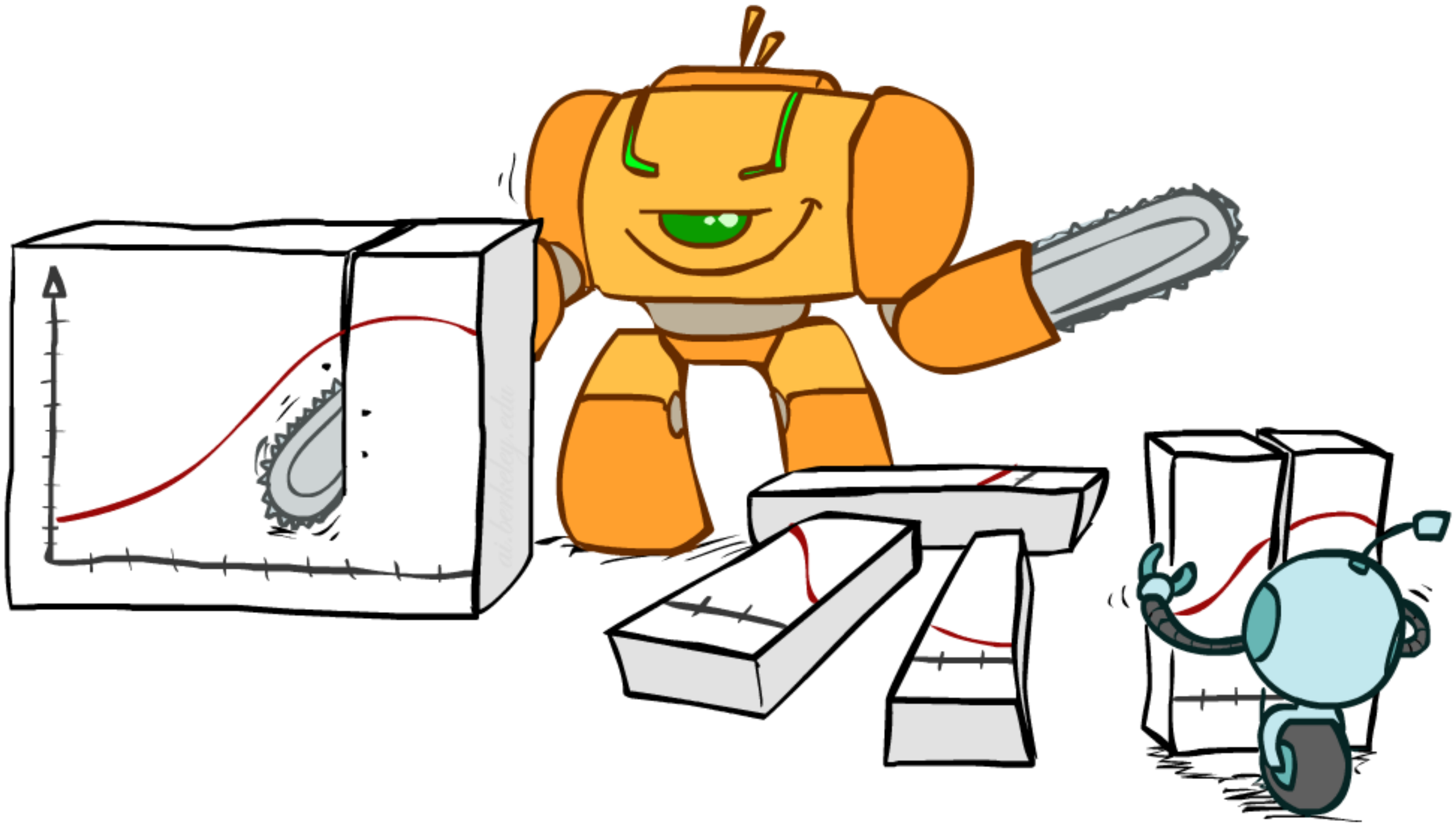
- Worst-Case Time Complexity: $O(d^n)$
- Space Complexity: $O(d^n)$ to store joint distribution
- Need $O(d^n)$ data points to estimate the entries in the joint distribution

Questions?

live and on sli.do #cse473

A Better Approach?

CONTEXT



Bayes' Rule

DEFINITION

Formulate conditional probability another way…

- From the product rule:
  - $P(A, B) = P(A|B) \cdot P(B)$
  - $P(A, B) = P(B|A) \cdot P(A)$
- Combine and rearrange:

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

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

Implications of Bayes' Rule

DEFINITION

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

- How does this help?
  - Calculate a condition from its reverse, which may be easy to calculate
  - Describes update to **posterior** $P(A|B)$ given **prior** $P(A)$ and **likelihood** $P(B|A)$

One of the most important equations in all of artificial intelligence!

Inference with Bayes' Rule

EXAMPLE

- Diagnostic probability from causal probability

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause}) \cdot P(\text{cause})}{P(\text{effect})}$$

- For example:
  - M is meningitis, S is stiff neck
  - Given $P(m) = 0.0001$, $P(s) = 0.01$, $P(s|m) = 0.8$

$$P(m|s) = \frac{P(s|m) \cdot P(m)}{P(s)} = \frac{0.8 \cdot 0.0001}{0.01} = 0.008$$

  - Posterior probability is still small, but 80x higher given evidence of stiff neck

Independence

FOUNDATIONS

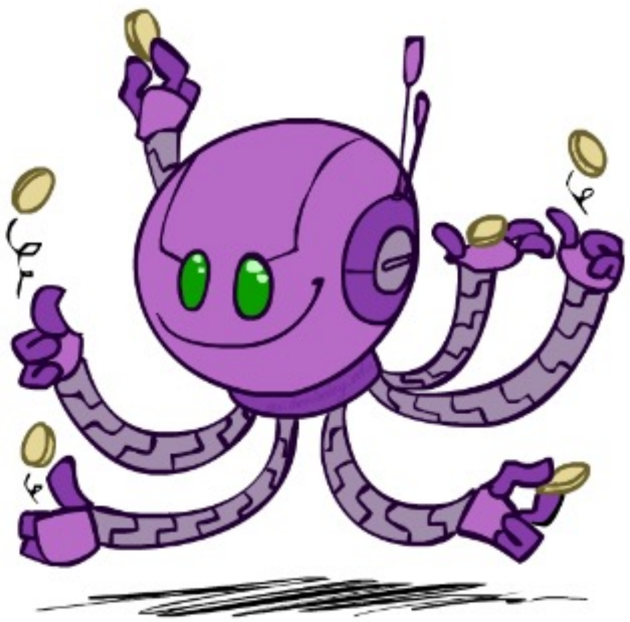
Two variables are independent if and only if:

$$\forall x, y \; P(x, y) = P(x) \cdot P(y) \Longrightarrow X \perp Y$$

Equivalently:

- The probability of x is not changed by whether y has occurred

$$P(x|y) = P(x)$$



- Example:
  - n fair, independent coin flips $X_1, \dots, X_n$
  - Possible 'worlds': 2^n

Conditional Independence

FOUNDATIONS

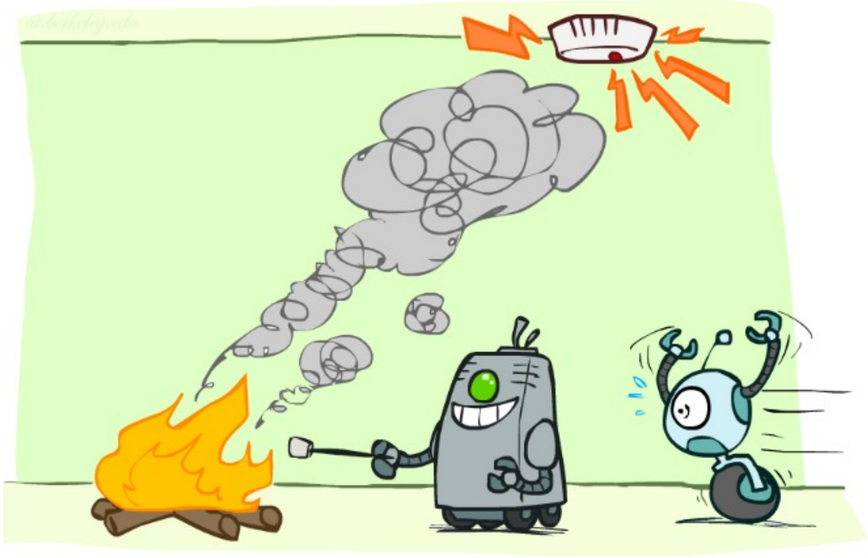
If two variables are ***conditionally* independent**, observing one won't influence our prediction of the other, provided we already know a third piece of information

- Formally:
  - X is conditionally independent of Y given Z if and only if:

$$\forall x, y, z \quad P(x|y,z) = P(x|z) \Longrightarrow X \perp Y \mid Z$$

  - Equivalently:

$$\forall x, y, z \quad P(x,y|z) = P(x|z) \cdot P(y|z)$$



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

Questions?

live and on sli.do #cse473

that's it for today!

SUMMARY

- Probability Review
- Rules of Probability
- Inference by Enumeration
- Bayes' Rule

UPCOMING

- Modeling with Probabilities

REMINDERS

- Complete **Practice Problem 10**
- Do **Homework 2** (due 7.23)
- Do **Project 3** (due 7.30)