CSPB3202 Artificial Intelligence

Reasoning with Uncertainty

University of Colorado Boulder

Note: Slides were adapted from the materials from Prof. Pieter Abbeel and Prof. Dan Klein at UC Berkeley, and Russell & Norvig's AI book.

Probability Review

Why we need probability

Probabilistic Reasoning vs. Logic

how much
degree

Cause and effect

Toothpain ⇒ Cavity 80%
implies

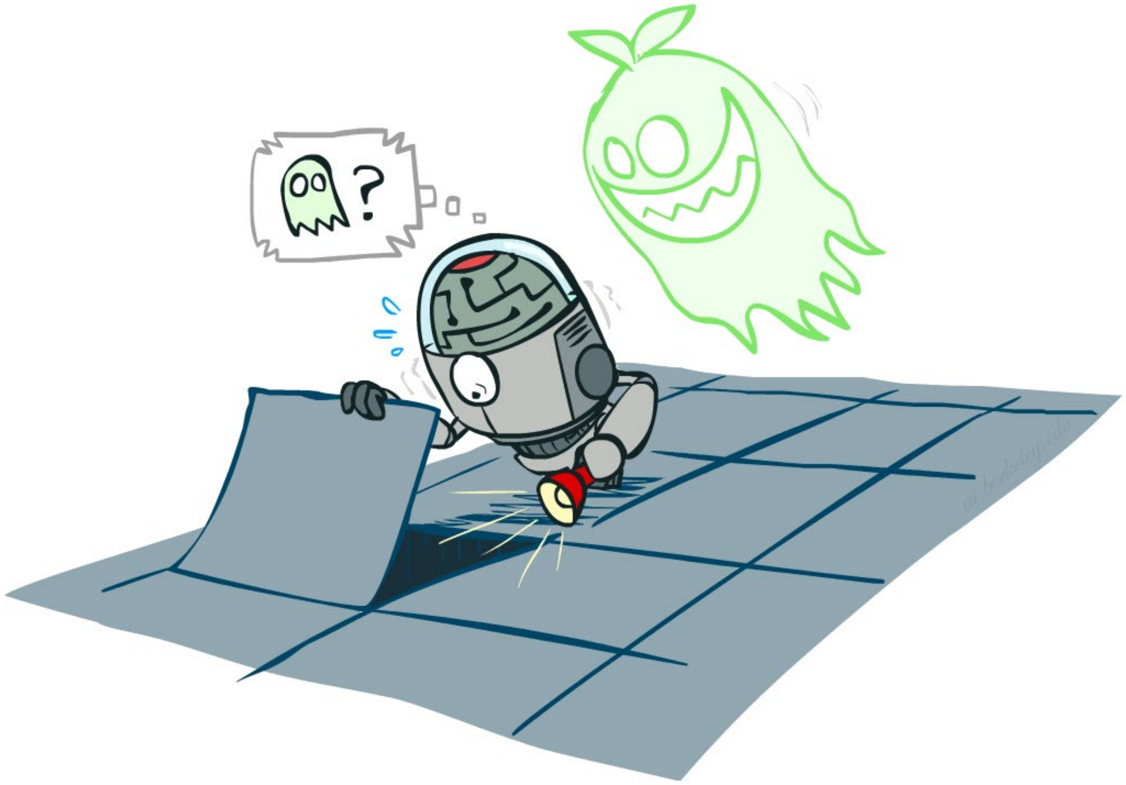
⇓ ⇓ Gum Disease 15%

fractures, cracks 5%

Why we need probability

Probabilistic Reasoning

- Diagnosis
- Speech recognition
- Tracking objects
- Robot mapping
- Genetics
- Error correcting codes
- … lots more!



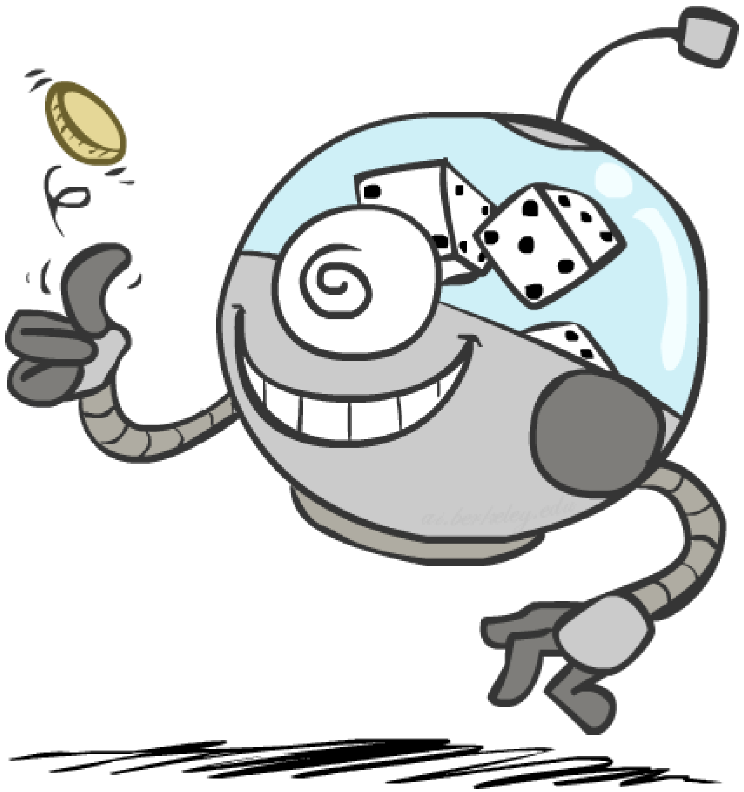
Probability

Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?

- We denote random variables with capital letters

- Like variables in a CSP, random variables have domains
  - R in {true, false}   (often write as {+r, -r})
  - T in {hot, cold}
  - D in [0, ∞)
  - L in possible locations, maybe {(0,0), (0,1), …}

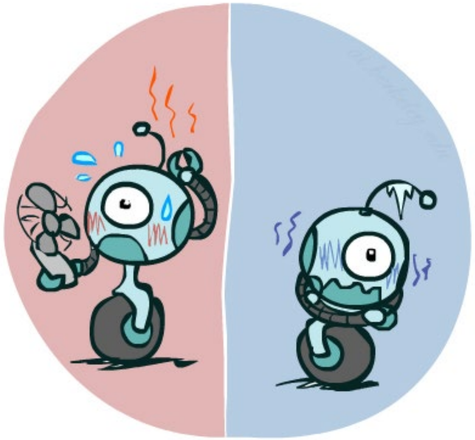


Probability Distributions

- Associate a probability with each value

  - Temperature:

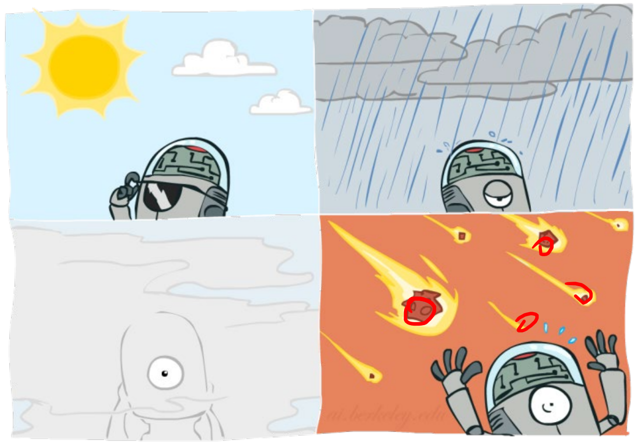
$(P_1+P_2+P_3+P_4) = 1$



$P(T)$

T	P
hot	0.5
cold	0.5

  - Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
~~fog~~ snow	0.29
~~meteor~~	0.01

hail

Probability Distributions

- Unobserved random variables have distributions

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Shorthand notation:

$$P(hot) = P(T = hot),$$
$$P(cold) = P(T = cold),$$
$$P(rain) = P(W = rain),$$
$$\ldots$$

OK *if* all domain entries are unique

- A distribution is a TABLE of probabilities of values

- A probability (lower case value) is a single number

- Must have: $P(W = rain) = 0.1$ and

$$\forall x \;\; P(X = x) \geq 0 \qquad\qquad \sum_x P(X = x) = 1$$

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$ specifies a real number for each assignment (or *outcome*):

$P(X_1 = a, X_2 = b)$

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$

$$P(x_1, x_2, \ldots x_n)$$

  - Must obey:

$$P(x_1, x_2, \ldots x_n) \geq 0$$

$$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

- Size of distribution if n variables with domain sizes d?
  - For all but the smallest distributions, impractical to write out!

d^n

Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny?
  - Probability that it's hot?
  - Probability that it's hot OR sunny?

P(H,S)

P(H ∪ S)

- Typically, the events we care about are *partial assignments*, like P(T=hot)

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

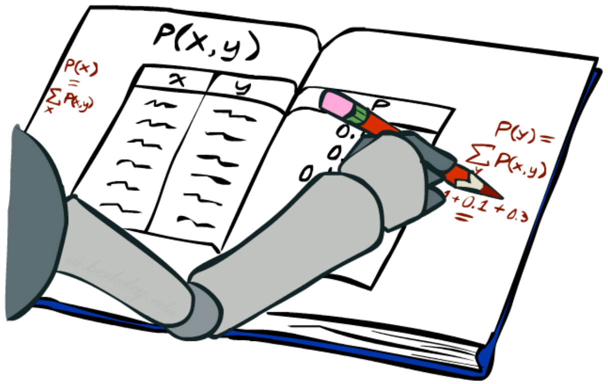
Quiz: Events

- P(+x, +y) ? 0.2
- P(+x) ? 0.5
- P(-y OR +x) ? 0.6

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions



- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(t) = \sum_s P(t, s)$$

$$P(s) = \sum_t P(t, s)$$

$$P(T)$$

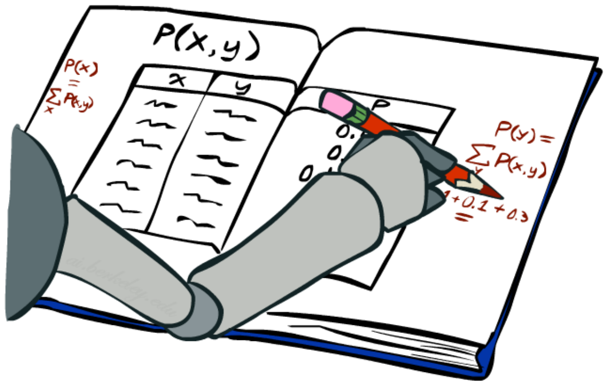
T	P
hot	0.5
cold	0.5

$$P(W)$$

W	P
sun	0.6
rain	0.4

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

Quiz: Marginal Distributions



$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x, y)$$

$$P(y) = \sum_x P(x, y)$$

$P(X)$

X	P
+x	0.5
-x	0.5

$P(Y)$

Y	P
+y	0.6
-y	0.4

Conditional Probabilities

- A simple relation between joint and conditional probabilities
  - In fact, this is taken as the *definition* of a conditional probability

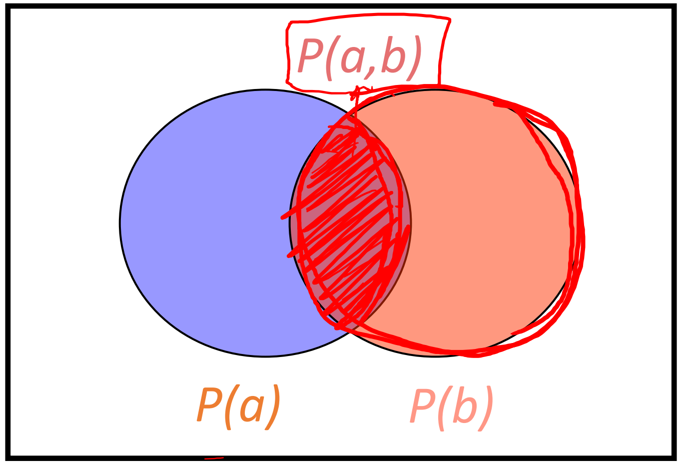
$P(a,b) = P(a \cap b)$

outcome

evidence (know)

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

not know



$S/t =$

$$P(T,W)$$

T	W	P
~~hot~~	~~sun~~	~~0.4~~
~~hot~~	~~rain~~	~~0.1~~
cold	sun	0.2
cold	rain	0.3

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)} = \frac{0.2}{0.5} = \boxed{0.4}$$

$$= P(W=s,T=c) + P(W=r,T=c)$$

$$= 0.2 + 0.3 \quad = 0.5$$

Quiz: Conditional Probabilities

$$P(X,Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

- P(+x | +y) ? $\frac{0.2}{0.2+0.4} = 1/3$
- P(-x | +y) ? $1 - 1/3 = 2/3$
- P(-y | +x) ? $\frac{0.3}{0.5} = 0.6$

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$$P(W|T = hot)$$

W	P
sun	0.8
rain	0.2

$$P(W|T = cold)$$

W	P
sun	0.4
rain	0.6

Joint Distribution

$$P(T, W)$$

T	W	P
~~hot~~	~~sun~~	~~0.4~~
~~hot~~	~~rain~~	~~0.1~~
cold	sun	0.2
cold	rain	0.3

Normalization Trick

$$P(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W = s|T = c) = \frac{P(W = s, T = c)}{\boxed{P(T = c)}}$$

$$= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)}$$

$$= \frac{0.2}{0.2 + 0.3} = 0.4$$

$$P(W = r|T = c) = \frac{P(W = r, T = c)}{P(T = c)}$$

$$= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)}$$

$$= \frac{0.3}{0.2 + 0.3} = 0.6$$

$$P(W|T = c)$$

W	P
sun	0.4
rain	0.6

Normalization Trick

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence

$P(c, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

Normalization Trick

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence

$P(c, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

- Why does this work? Sum of selection is P(evidence)! (P(T=c), here)

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{P(x_1, x_2)}{\sum_{x_1} P(x_1, x_2)}$$

Quiz: Normalization Trick

- P(X | Y=-y) ?

$$P(X, Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

SELECT the joint probabilities matching the evidence →

+x , -y | 0.3 ~~0.4~~
-x , -y | 0.1 ~~0.4~~

NORMALIZE the selection (make it sum to one) →

X	Y	P
+x	-y	0.75
-x	-y	0.25

Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$ known.
  - Query* variable: Q Interested
  - Hidden variables: $H_1 \dots H_r$ unobserved

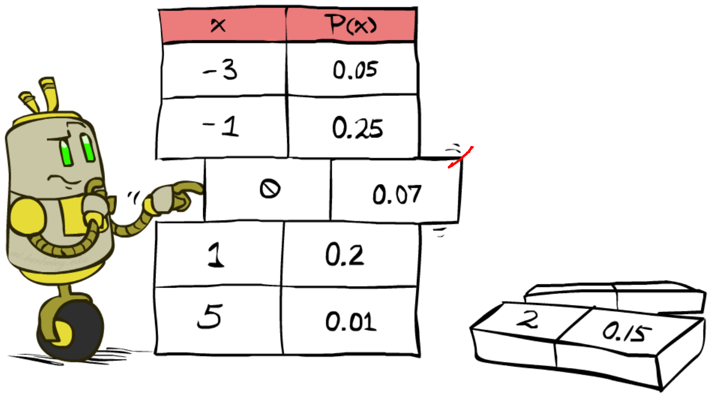
$\left.\right\}$ $X_1, X_2, \dots X_n$ *All variables*

- We want:

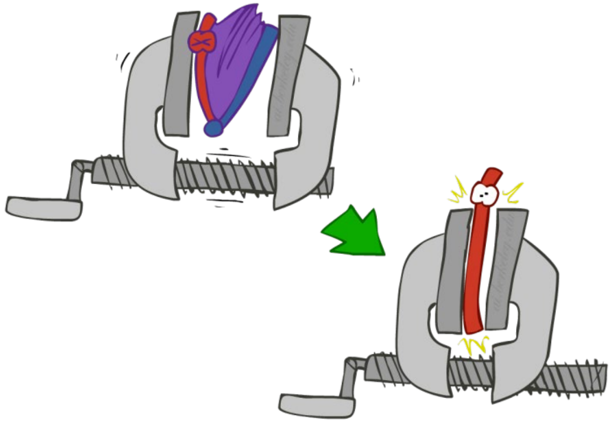
* *Works fine with multiple query variables, too*

$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Inference by Enumeration

- P(W)?

Q → weather < sun, rain

0.3 + 0.1 + 0.1 + 0.15 = 0.65

= 1 − 0.65 = 0.35

E → None

H → S, T

- P(W | winter)?

s 0.5

r 0.5

$\frac{0.25}{0.5} = 0.5$

- P(W | winter, hot)?

s | w, h = $\frac{0.1}{0.15} = 2/3$

r | w, h = 1 − 2/3 = $\frac{0.05}{0.15}$ = 1/3

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Z

Inference by Enumeration

- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

The Product Rule

- Sometimes have conditional distributions but want the joint

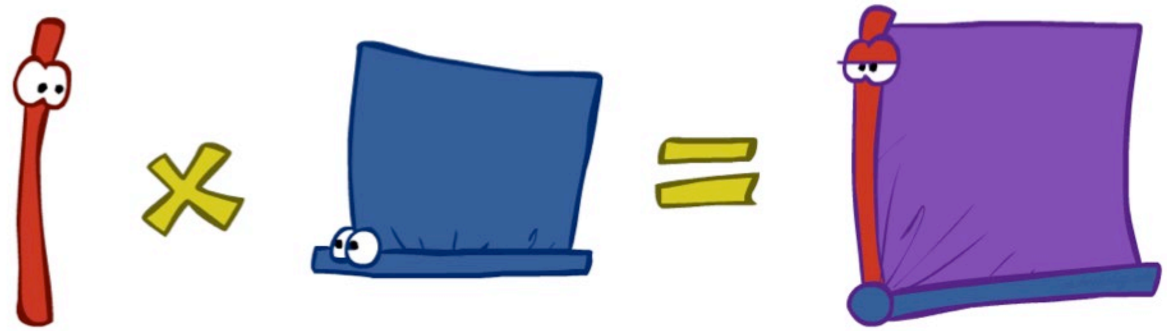
posterior

conditional   Joint

$$P(y)P(x|y) = P(x,y) \quad \Longleftrightarrow \quad P(x|y) = \frac{P(x,y)}{P(y)}$$

prior

known



The Product Rule

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

- Example:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$P(D, W)$

D	W	P
wet	sun	0.1 × 0.8
dry	sun	0.9 × 0.8
wet	rain	
dry	rain	

The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions

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

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{N} P(x_i|x_1 \ldots x_{i-1})$$

- Why is this always true?

$P(x, y) = P(x|y) \cdot P(y)$

$P(x_1, x_2, x_3)$

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

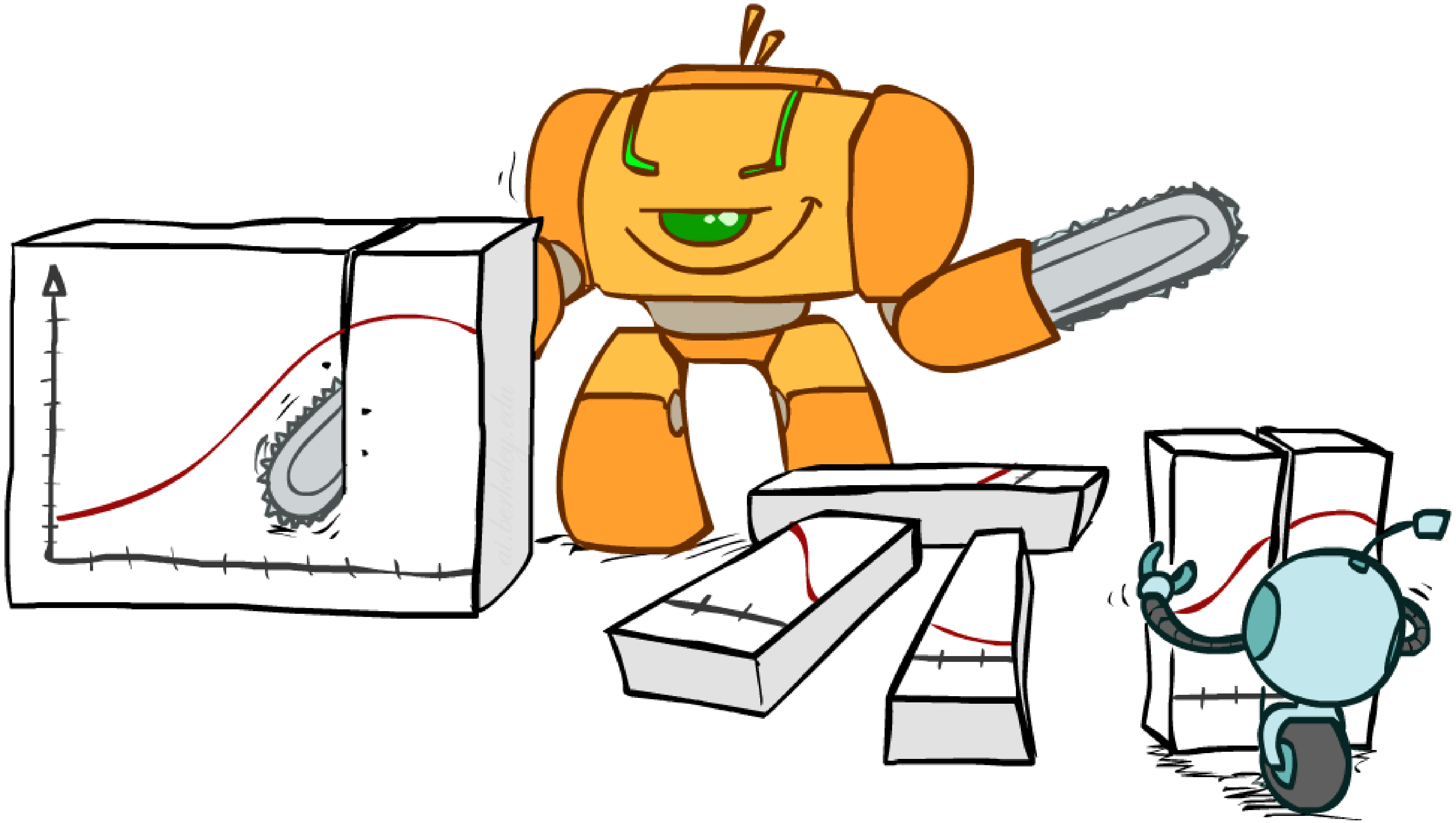
$P(x_2, x_3|x_1)$

$\boxed{P(x_2|x_1)} \, P(y)$

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

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

Bayes Rule



Bayes' Rule

- Two ways to factor a joint distribution over two variables:

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

- Dividing, we get:

$$\boxed{P(x|y) = \frac{P(y|x)}{P(y)}P(x)}$$

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple

- In the running for most important AI equation!

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

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

- Example:
  - M: meningitis, S: stiff neck

$$\left.\begin{aligned} P(+m) &= 0.0001 \\ \boxed{P(+s|+m)} &= 0.8 \\ P(+s|-m) &= 0.01 \end{aligned}\right\} \text{Example givens}$$

?

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{\boxed{P(+s)}} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \boxed{\frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}}$$

+s	-m
+s	+m

≈ 0.0079

0.8%

- Note: posterior probability of meningitis still very small
- Note: you should still get stiff necks checked out! Why?

Quiz: Bayes' Rule

- Given:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$\times 0.8$

$\times 0.2$

- What is P(W | dry) ?

$$P(W=s \mid dry) = \frac{\overset{0.9}{P(dry|sun)} \cdot \overset{0.8}{P(sun)}}{P(dry) = 0.72 + 0.06}$$

$$P(W=r \mid dry) = \frac{\overset{0.3}{P(dry|rain)} \cdot \overset{0.2}{P(rain)}}{P(dry) = 0.78}$$

Inference with Bayes' Rule

- Example: Breast Cancer Diagnosis from Mammogram Screening

$$P(c \mid +)$$

$$= \frac{\boxed{P(+ \mid c) \cdot P(c)}}{P(+)} \rightarrow P(+, c)$$

$$= \frac{0.009}{0.009 + 0.05 \times 0.99}$$

$$= 0.15$$

15%

$P(C) = 1\%$ 40–50

$TPR = 90\% \rightarrow P(+ \mid C) = 0.9$

$FPR = 5\% \rightarrow P(+ \mid NC) = 0.05$

C/NC	+/−	P(C,+)
C	+	0.009
C	−	0.001
NC	+	
NC	−	

Joint prob

$P(C) \cdot P(+ \mid C)$

0.01×0.9

$P(C) \cdot 1 - P(+ \mid C) = 0.001$

$P(+ \mid NC) \cdot P(NC) = 0.05 \cdot 0.99$

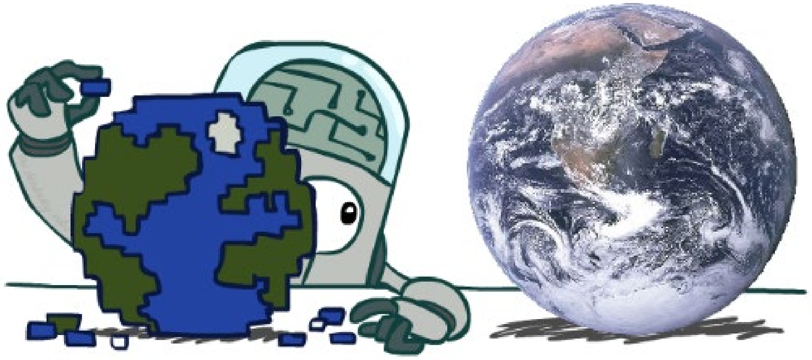
$0.95 \cdot 0.99$

Probabilistic Models

- Models describe how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information

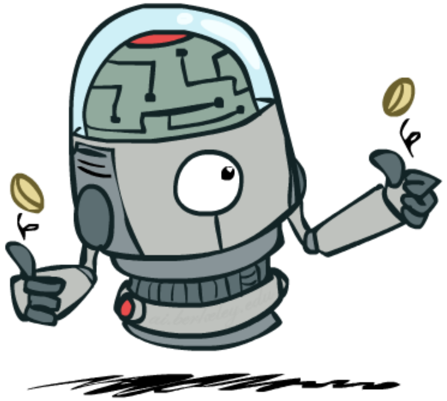


Independence

- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form: $\forall x, y : P(x|y) = P(x)$
  - We write: $\boxed{X \perp\!\!\!\perp Y}$



- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

×

$P(X_2)$

H	0.5
T	0.5

× ⋯ ×

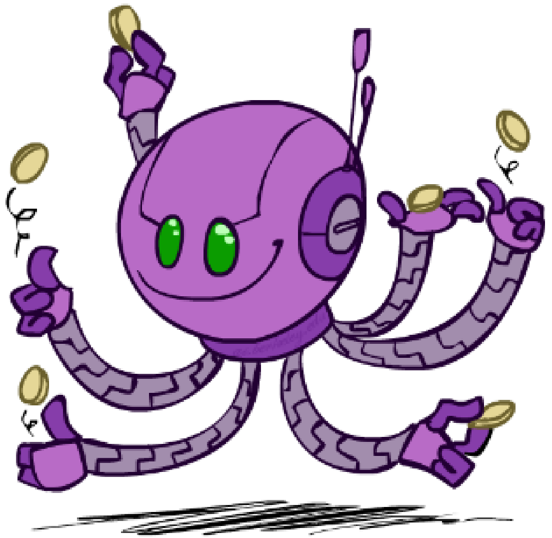
$P(X_n)$

H	0.5
T	0.5

$P(X_1, X_2, \ldots X_n)$

2^n

$P(X_1 = H | \ldots)$



Conditional Independence

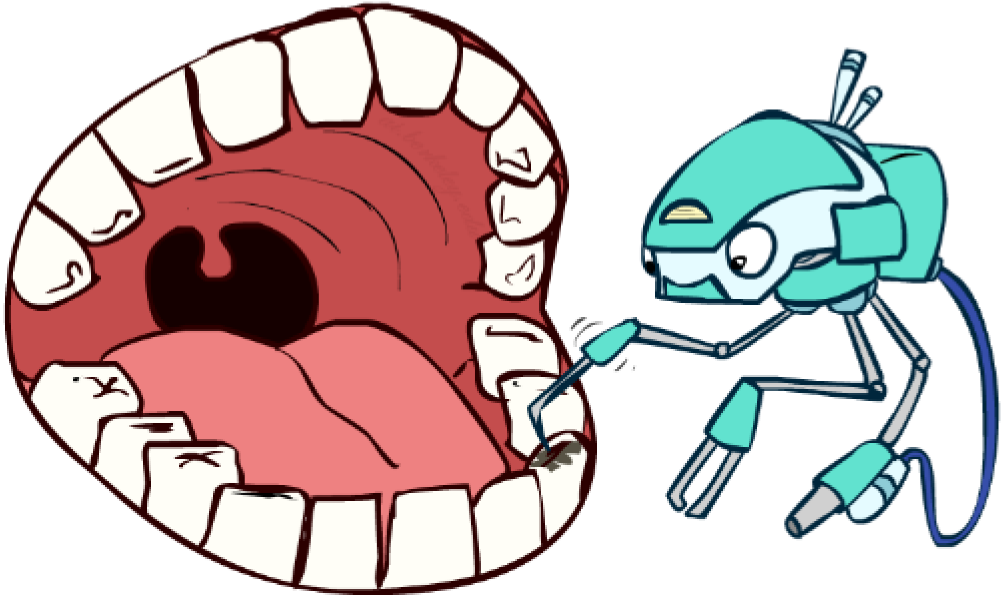
- P(Toothache, Cavity, Catch)

  +/- +/- +/-

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

- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare (why?)

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

- X is conditionally independent of Y given Z

$$\boxed{X \perp\!\!\!\perp Y | Z}$$

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

if and only if:

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

or, equivalently, if and only if

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

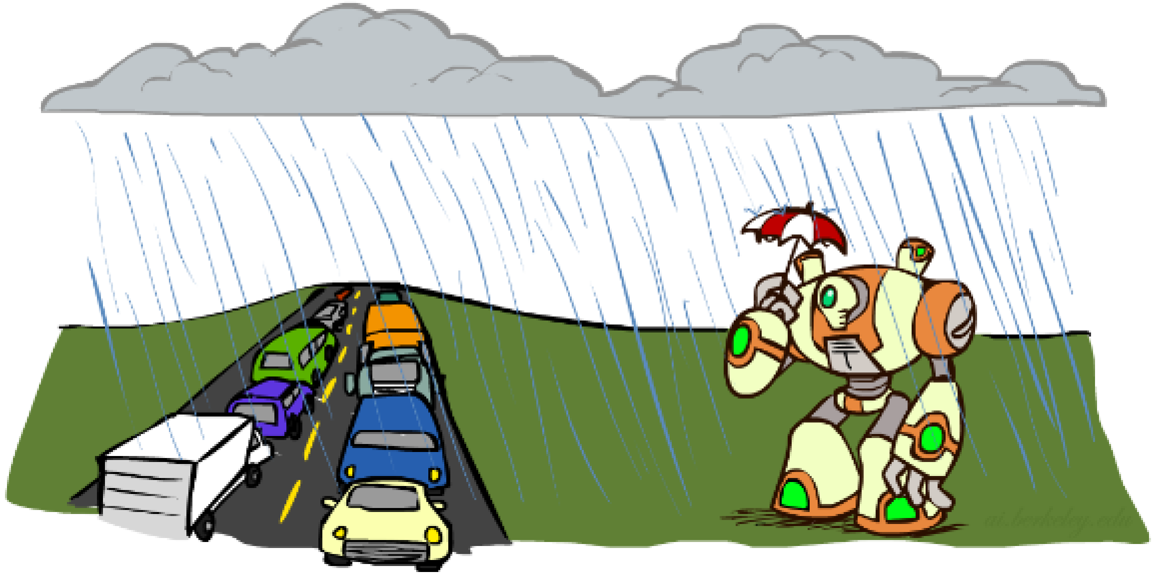
Conditional Independence

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

$R \to T$ ✓

$R \to U$ ✓

$T \perp\!\!\!\perp U \mid R$



Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm

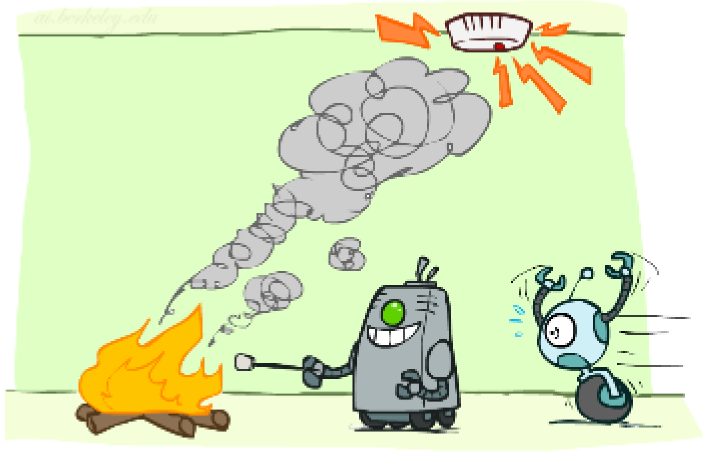
$T \leftarrow R \rightarrow U$

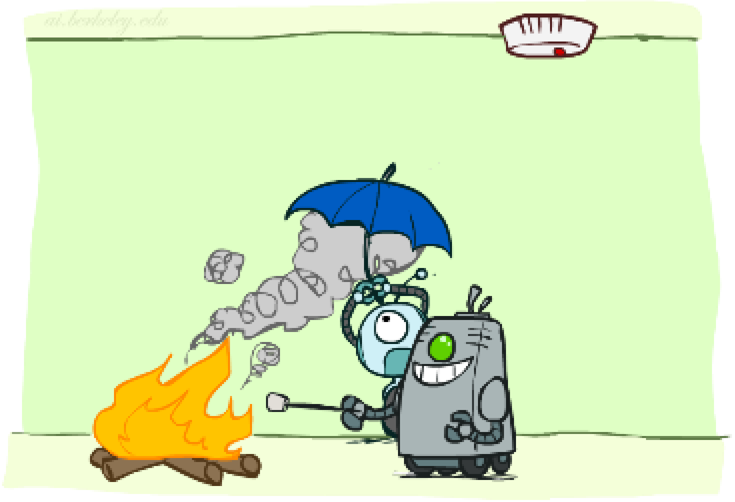
$F \rightarrow S \rightarrow A$

$F \rightarrow S$

$S \rightarrow A$

$F \perp\!\!\!\perp A \mid S$





Conditional Independence and the Chain Rule

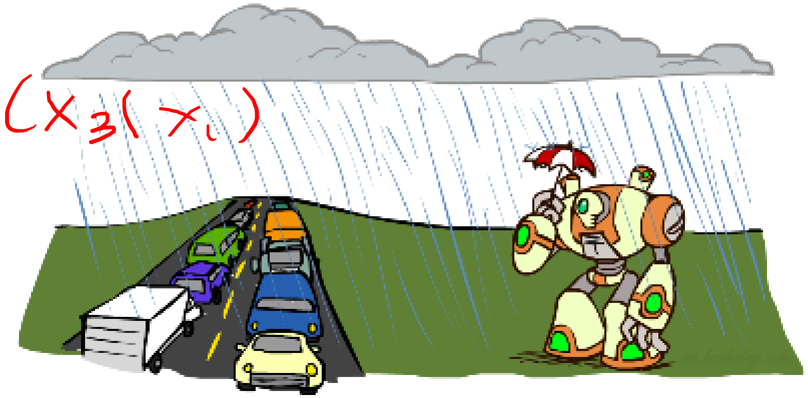
- Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

X_1, X_2

$P(X_3 | X_1)$

- Trivial decomposition:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \cancel{\text{Traffic}})$$



- With assumption of conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) =$$
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- Bayes'nets / graphical models help us express conditional independence assumptions