CSE 473: Artificial Intelligence

Probability

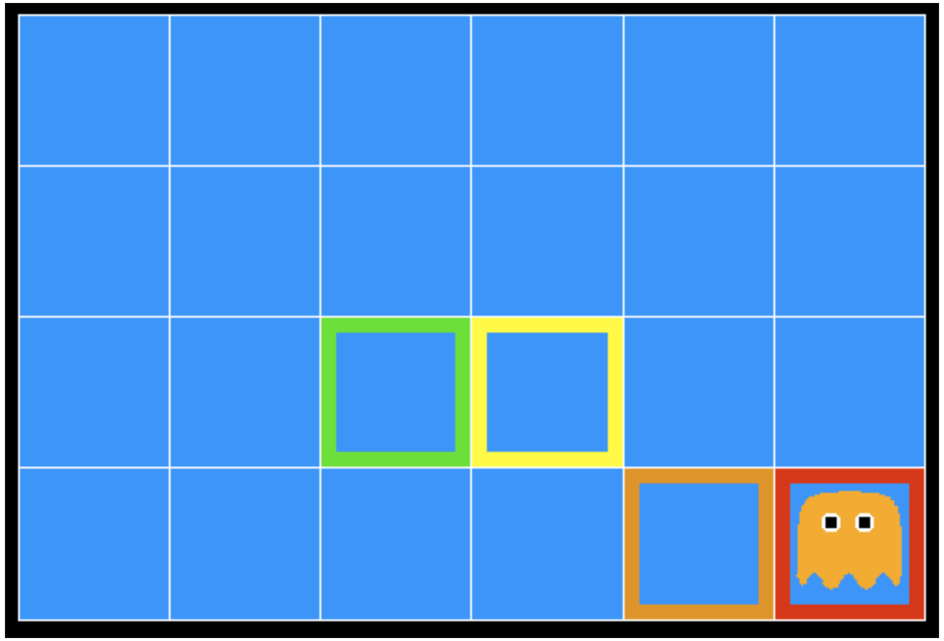
slides adapted from
Stuart Russel, Dan Klein, Pieter Abbeel from ai.berkeley.edu
And Hanna Hajishirzi, Jaren Moore, Dan Weld

Uncertainty

- The real world is rife with uncertainty!
  - E.g., if I leave for SEA 60 minutes before my flight, will arrive in time?
- Problems:
  - partial observability (road state, other drivers' plans, etc.)
  - noisy sensors (radio traffic reports, Google maps)
  - immense complexity of modelling and predicting traffic, security line, etc.
  - lack of knowledge of world dynamics (will tire burst? need COVID test?)
- Combine probability theory + utility theory -> decision theory
  - ***Maximize expected utility*** : $a^* = \mathrm{argmax}_a \sum_s P(s \mid a)\, U(s)$

Inference in Ghostbusters

- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
  - On the ghost: red
  - 1 or 2 away: orange
  - 3 or 4 away: yellow
  - 5+ away: green

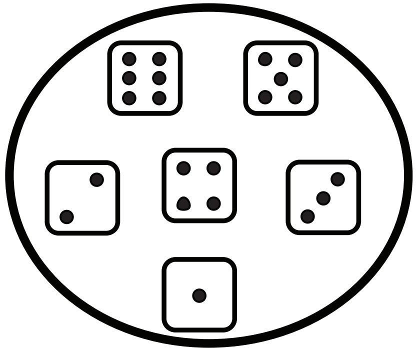


- Sensors are noisy, but we know P(Color(x,y) | DistanceFromGhost(x,y))

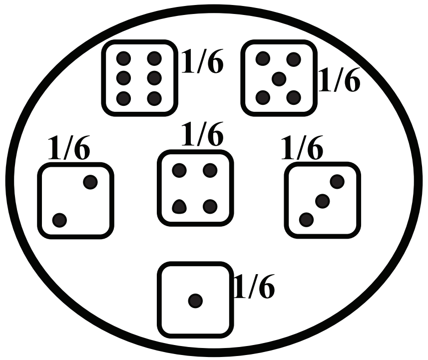
P(red | 3)	P(orange | 3)	P(yellow | 3)	P(green | 3)
0.05	0.15	0.5	0.3

Basic laws of probability

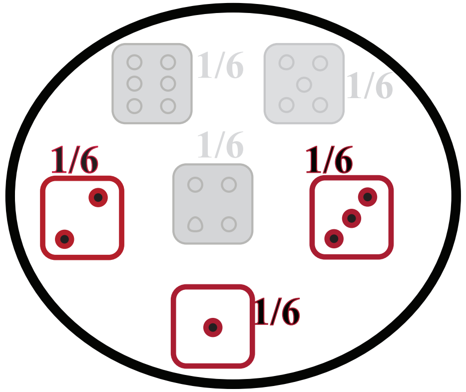
- Begin with a set Ω of possible worlds
  - E.g., 6 possible rolls of a die, {1, 2, 3, 4, 5, 6}

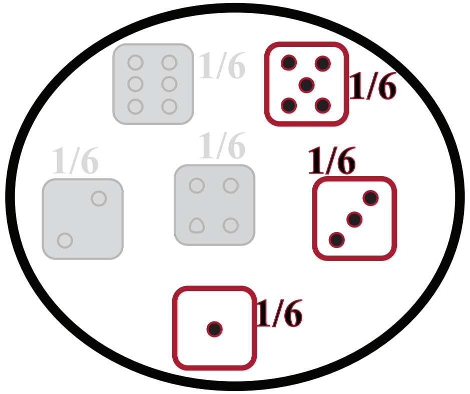


- A ***probability model*** assigns a number $P(\omega)$ to each world ω
  - E.g., $P(1) = P(2) = P(3) = P(5) = P(5) = P(6) = 1/6$.
- These numbers must satisfy
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$



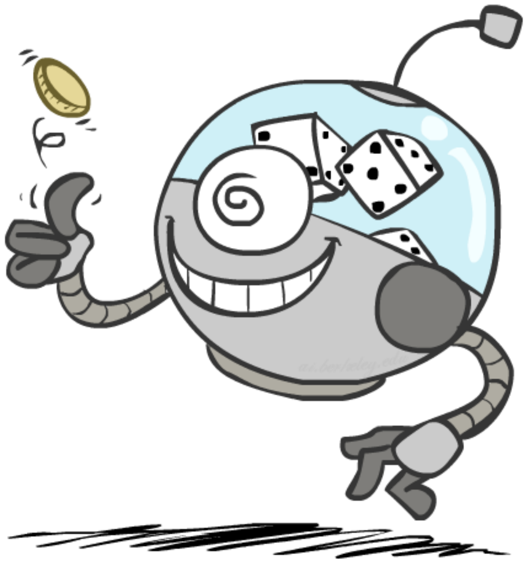
Basic laws contd.





- An ***event*** is any subset of Ω
  - E.g., “roll < 4” is the set {1,2,3}
  - E.g., “roll is odd” is the set {1,3,5}
- The probability of an event is the ***sum*** of probabilities over its worlds
  - $\boldsymbol{P}(A) = \sum_{\omega \in A} P(\omega)$
  - E.g., $P(roll < 4) = P(1) + P(2) + P(3) = 1/2$
- De Finetti (1931):
  - anyone who bets according to probabilities that violate these laws can be forced to lose money on every set of bets

Random Variables



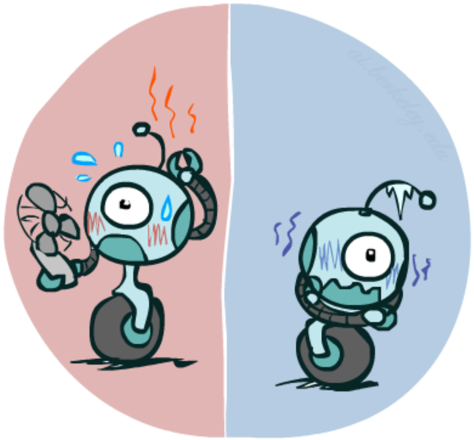
- A random variable (usually denoted by a capital letter) is some aspect of the world about which we (may) be uncertain
  - Formally a ***deterministic function*** of ω
- The ***range*** of a random variable is the set of possible values
  - *Odd* = Is the dice roll an odd number? $\rightarrow$ {true, false}
    - e.g. *Odd*(1)=true, *Odd*(6) = false
    - often write the event *Odd*=true as *odd*, *Odd*=false as $\neg$*odd*
  - *T* = Is it hot or cold? $\rightarrow$ {hot, cold}
  - *D* = How long will it take to get to the airport? $\rightarrow$ $[0, \infty)$
  - L_{Ghost} = Where is the ghost? $\rightarrow$ {(0,0), (0,1), …}
- The ***probability distribution*** of a random variable *X* gives the probability for each value *x* in its range (probability of the event *X*=*x*)
  - $P(X=x) = \sum_{\{\omega:\, X(\omega)=x\}} P(\omega)$
  - *P*(*x*) for short (when unambiguous)
  - $\mathbf{P}(X)$ refers to the entire distribution (think of it as a vector or table)

Probability Distributions

- Associate a probability with each value; sums to 1
  - Temperature:
  - Weather:
  - ***Joint distribution***

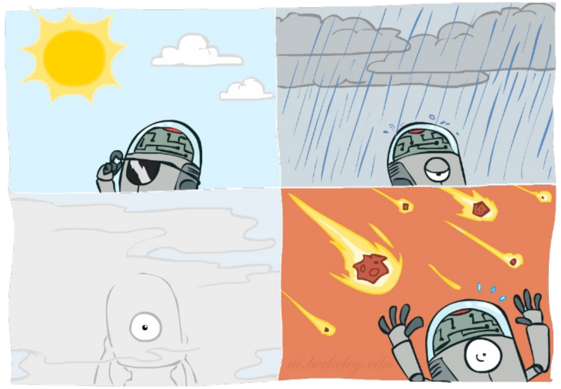
P(T)

T	P
hot	0.5
cold	0.5



P(W)

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0



P(T,W)

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

Making possible worlds

- In many cases we
  - begin with random variables and their domains
  - construct possible worlds as assignments of values to all variables
- E.g., two dice rolls $Roll_1$ and $Roll_2$
  - How many possible worlds?
  - What are their probabilities?
- Size of distribution for n variables with range size d? d^n
  - For all but the smallest distributions, cannot write out by hand!

Probabilities of events

- The Probability of an event is the sum of probabilities of its worlds, $P(A) = \sum_{\omega \in A} P(\omega)$
- So, given a joint distribution over all variables, can compute any event probability!
  - Probability that it's hot AND sunny?
    - $P(T=hot, W=sun)$
    - $= .45$
  - Probability that it's hot?
    - $P(T=hot) = \sum_{w \in W} P(T=hot, W=w)$
    - $= P(T=hot, W=sun) + P(T=hot, W=rain) + P(T=hot, W=fog) + P(T=hot, W=meteor)$
    - $= .45 + .02 + .03 + .00 = .5$
  - Probability that it's hot OR not foggy?
    - $P(T=hot \lor \neg W=fog) = P(T=hot) + P(\neg W=fog) - P(T=hot, \neg W=fog)$
    - $= P(T=hot) + (1 - P(W=fog)) - P(T=hot, \neg W=fog)$
    - $= .5 + (1 - .03 + .27) – (.45 + .02 + .00) = .5 + .7 - .47 = .73$

Joint distribution

P(T,W)

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

Quiz: Events

- P(+x, +y) ?
- P(+x) ?
- P(-y OR +x) ?

$$P(X, Y)$$

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

Quiz: Events

- P(+x, +y) ?

  = .2

- P(+x) ?

  = .2 + .3 = .5

- P(-y OR +x) ?

  = P(-y) + P(+x) - P(-y, + x) = .3 + .1 + .2 + .3 - .3 = .6
  = 1 - P(+y, -x) = 1 - .4 = .6

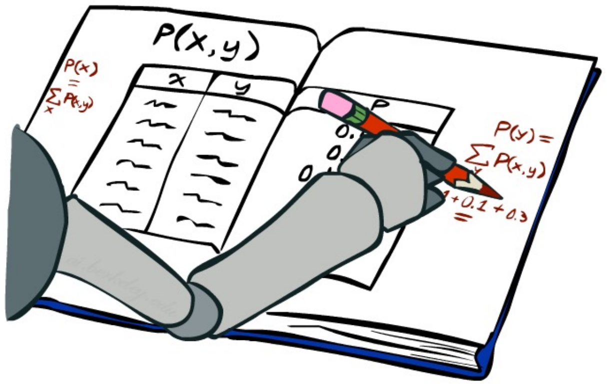
$P(X, Y)$

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

Marginal Distributions

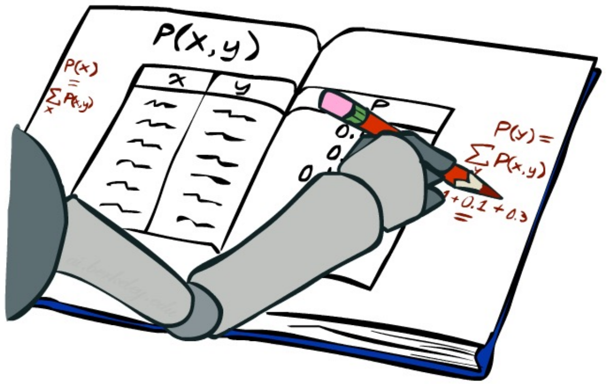
- Marginal distributions are sub-tables which eliminate variables
- ***Marginalization*** (***summing out***): Collapse a dimension by adding

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



		Temperature			
		hot	cold		
Weather	sun	0.45	0.15	0.60	**P**(W)
	rain	0.02	0.08	0.10	
	fog	0.03	0.27	0.30	
	meteor	0.00	0.00	0.00	
		0.50	0.50		
		P(T)			

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

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

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

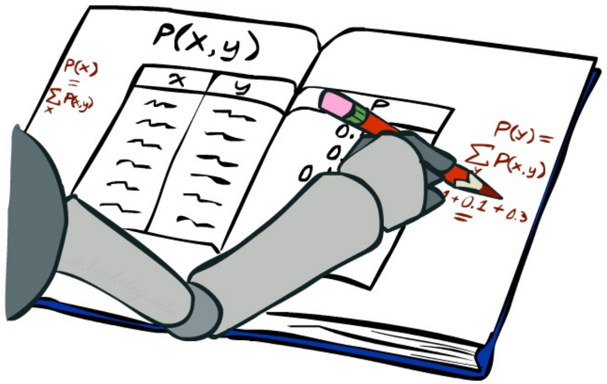
$$P(X)$$

X	P
+x	
-x	

$$P(Y)$$

Y	P
+y	
-y	

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

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

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

$P(X)$

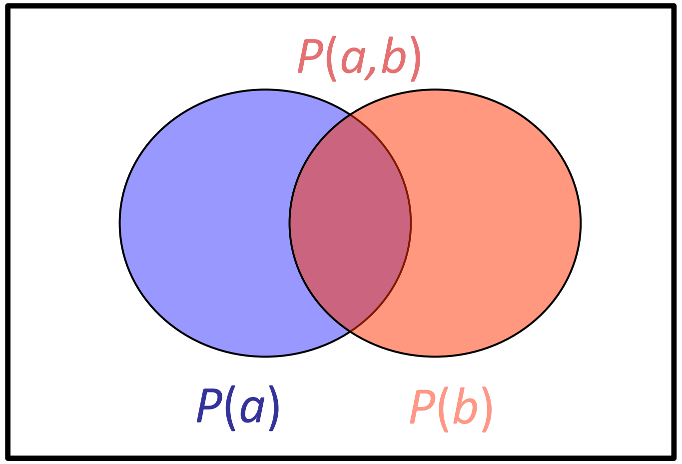
X	P
+x	.5
-x	.5

$P(Y)$

Y	P
+y	.6
-y	.4

Conditional Probabilities

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



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

P(T,W)

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

$$P(W=s \mid T=c) = \frac{P(W=s,T=c)}{\mathbf{P}(T=c)} = 0.15/0.50 = 0.3$$

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

$= 0.15 + 0.08 + 0.27 + 0.00 = 0.50$

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) ?

- P(-x \| +y) ?

- P(-y \| +x) ?

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) ?

  = .2 / (.2 + .4) = 1/3

- P(-x | +y) ?

  = .4 / (.2 + .4) = 2/3

- P(-y | +x) ?

  = .3 / (.3 + .2) = .6

Conditional Distributions

- Distributions for one set of variables given another set

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

P(W | T=h)

hot
0.90
0.04
0.06
0.00

P(W | T=c)

cold
0.30
0.16
0.54
0.00

P(W | T)

hot	cold
$\boxed{0.90}$	$\boxed{0.30}$
0.04	0.16
0.06	0.54
0.00	0.00

Notice how the values in the tables have been re-normalized!

Normalizing a distribution

- Procedure:
  - Multiply each entry by α = 1/(sum over all entries)

Ensure entries sum to ONE

P(W,T)

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

P(W,T=c)

0.15
0.08
0.27
0.00

$P(W \mid T=c) = P(W,T=c)/P(T=c)$

$= \alpha\ P(W,T=c)$

Normalize

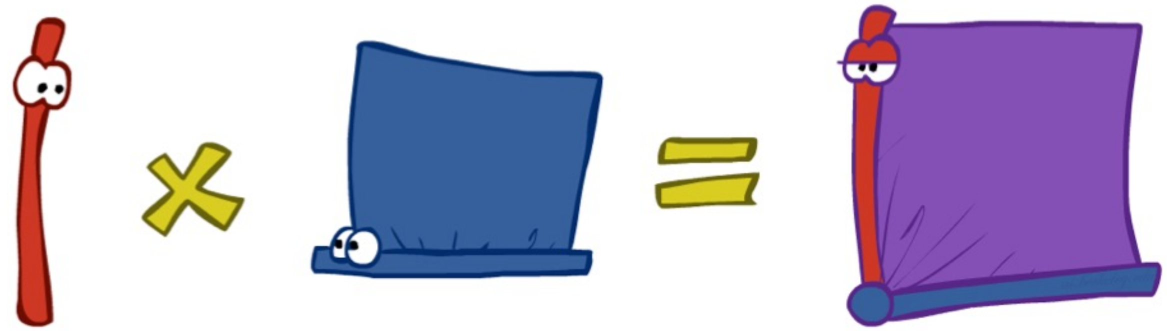
$\alpha = 1/0.50 = 2$

0.30
0.16
0.54
0.00

The Product Rule

- Sometimes we have conditional distributions but we want the joint

$$P(a \mid b)\, P(b) = P(a, b) \quad \Longleftrightarrow \quad P(a \mid b) = \frac{P(a, b)}{P(b)}$$



The Product Rule: Example

$$\mathbf{P}(W \mid T)\,\mathbf{P}(T) = \mathbf{P}(W, T)$$

P(W | T)

hot	cold
0.90	0.30
0.04	0.16
0.06	0.54
0.00	0.00

P(T)

T	P
hot	0.5
cold	0.5

P(W, T)

		Temperature	
		hot	cold
Weather	sun	0.45	0.15
	rain	0.02	0.08
	fog	0.03	0.27
	meteor	0.00	0.00

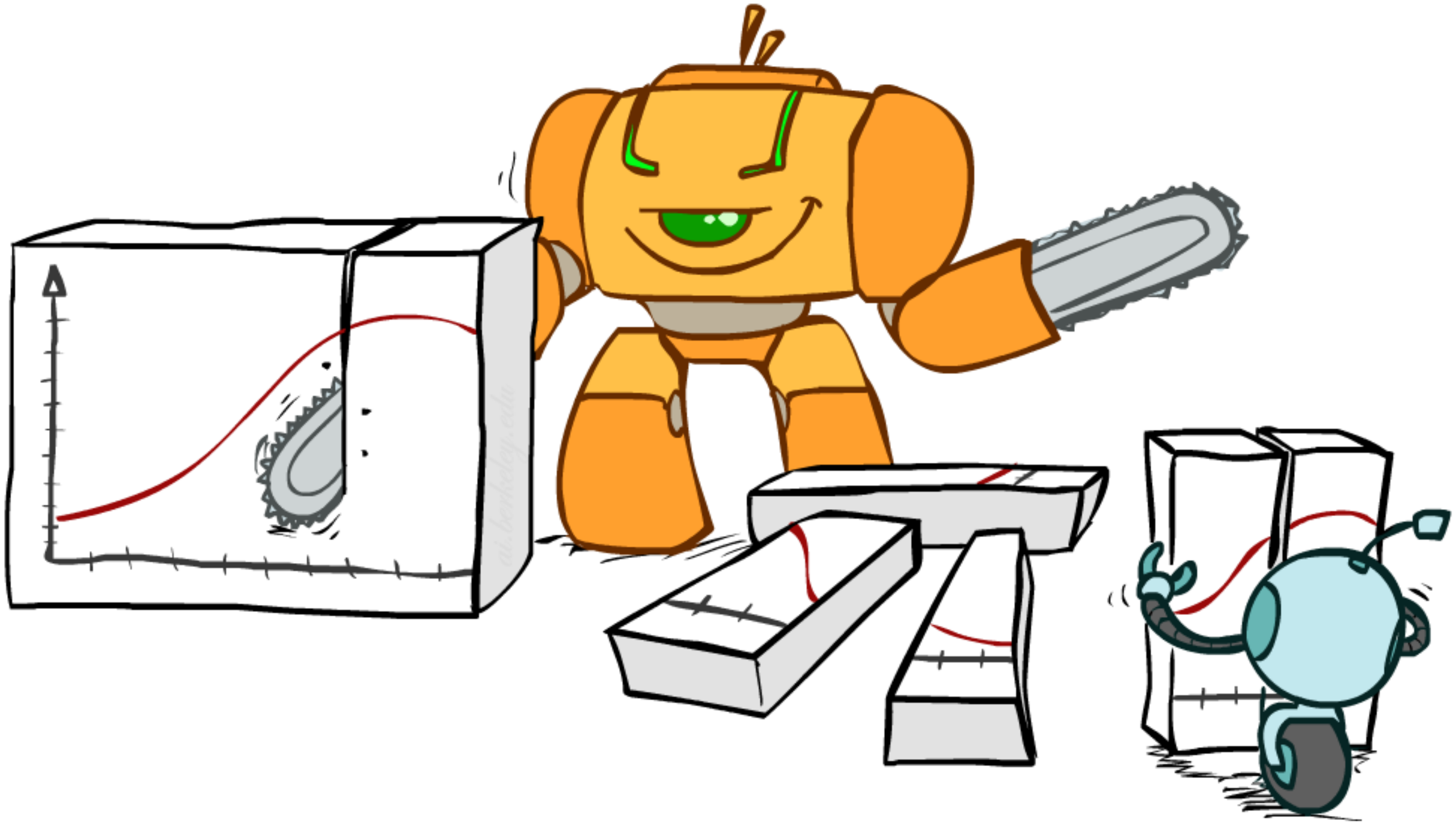
The Chain Rule

- A joint distribution can be written as a product of conditional distributions by repeated application of the product rule:

$$
\begin{aligned}
P(x_1, x_2, x_3) &= P(x_3 \mid x_1, x_2)\, P(x_1, x_2) \\
&= P(x_3 \mid x_1, x_2)\, P(x_2 \mid x_1)\, P(x_1)
\end{aligned}
$$

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

Bayes' Rule



Bayes' Rule

- Write the product rule both ways:

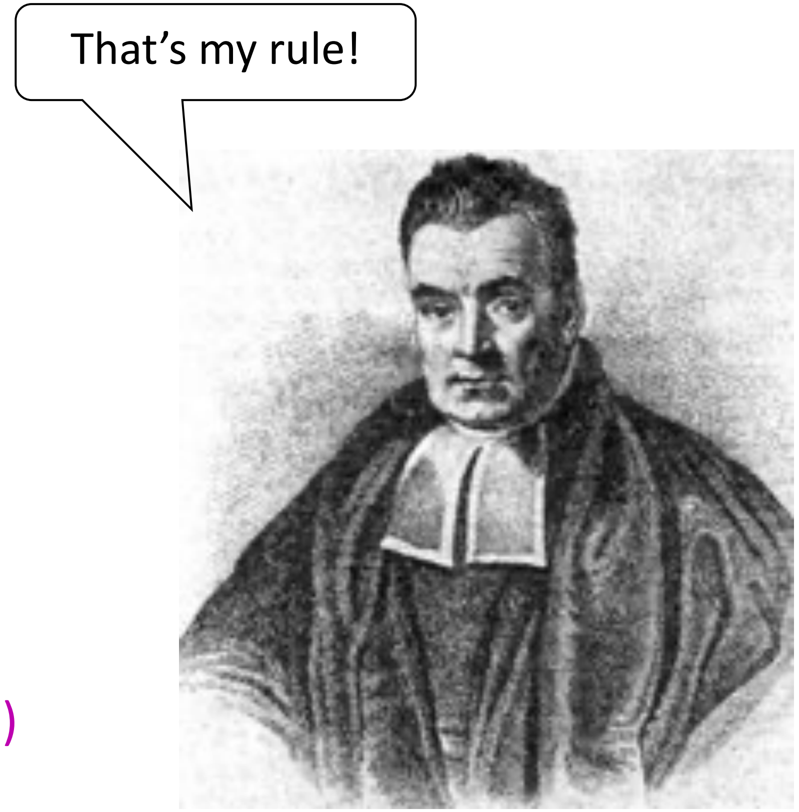
$$P(a \mid b)\, P(b) = P(a, b) = P(b \mid a)\, P(a)$$

- Dividing left and right expressions, we get:

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

- 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
  - Describes an "update" step from prior $P(a)$ to posterior $P(a \mid b)$
  - Foundation of many systems we'll see later

- In the running for most important AI equation!



Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

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

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

$$\left.\begin{array}{l} P(s \mid m) = 0.8 \\ P(m) = 0.0001 \\ P(s) = 0.01 \end{array}\right\} \text{Example givens}$$

$$P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.01}$$

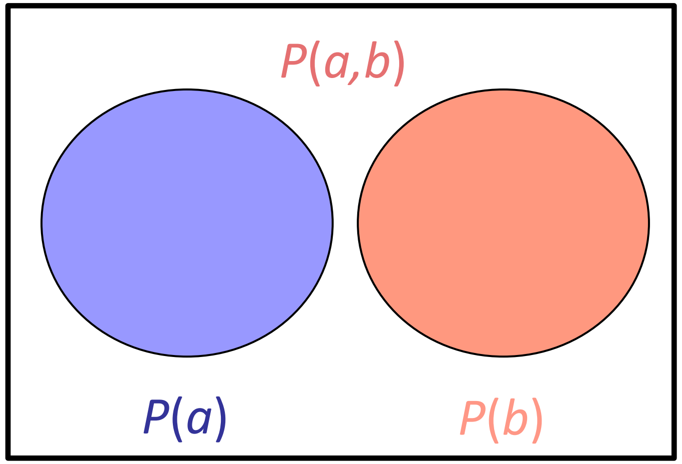
  - Note: posterior probability of meningitis still very small: 0.008 (80x bigger – why?)
  - Note: you should still get stiff necks checked out! Why?

Independence

- Two variables X and Y are (absolutely) ***independent*** if

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

  - I.e., the joint distribution ***factors*** into a product of two simpler distributions



- Equivalently, via the product rule $P(x,y) = P(x|y)P(y)$,

$$P(x \mid y) = P(x) \quad \text{or} \quad P(y \mid x) = P(y)$$

- Example: two dice rolls $Roll_1$ and $Roll_2$
  - $P(Roll_1{=}5, Roll_2{=}3) \;=\; P(Roll_1{=}5)\, P(Roll_2{=}3) = 1/6 \times 1/6 \;=\; 1/36$
  - $P(Roll_2{=}3 \mid Roll_1{=}5) \;=\; P(Roll_2{=}3)$

Example: Independence

- *n* fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

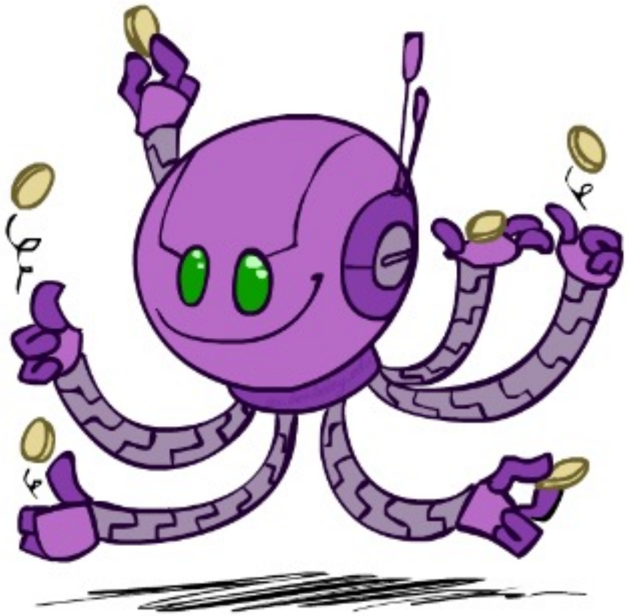
$P(X_2)$

H	0.5
T	0.5

. . .

$P(X_n)$

H	0.5
T	0.5



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

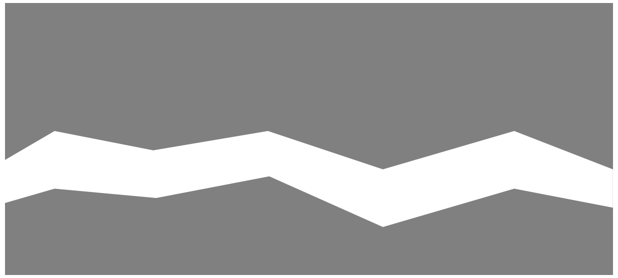
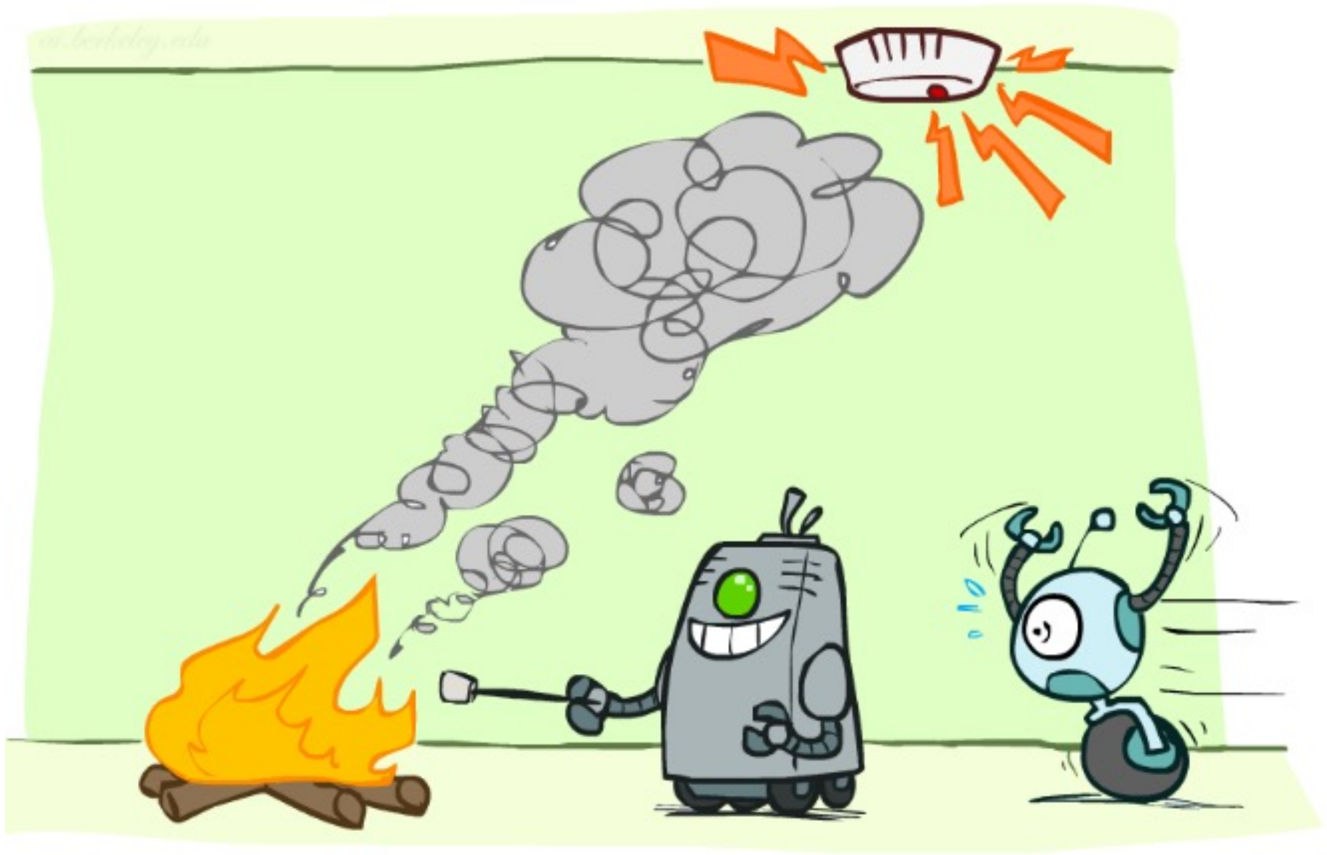


table size: 2^n

in general: d^n

Conditional Independence



Conditional Independence

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

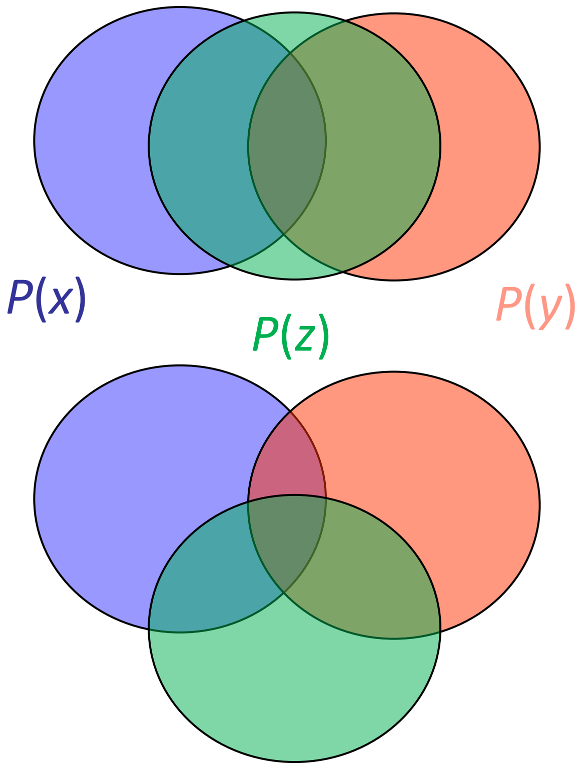
- X is conditionally independent of Y given Z :

$$\forall x,y,z \qquad P(x \mid y, z) = P(x \mid z)$$

$$= P(x,y,z) \,/\, P(y, z) = P(x,z) / P(z)$$

or, equivalently, if and only if

$$\forall x,y,z \qquad P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$$



Probabilistic Inference

- Probabilistic inference: compute a desired probability from a probability model
  - Typically for a ***query variable*** given ***evidence***
  - E.g., P(airport on time | no accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

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