Announcements

- Midterm 1
  - is on Monday, 7/15, during lecture time (12:30 – 2 pm in Dwinelle 155).
  - Mesut and Arin will be holding a MT1 review session 7 – 9 pm on Thursday, in Cory 521.
  - We will be releasing a ‘midterm prep page’ on the website which has all the information you need to know.
- Please tag your project partner when you submit to Gradescope. Otherwise, we will not be able to give them a score. Do this ASAP.
- HW3 due on Friday, 7/12, at 11:59 pm
- P2 due on Friday, 7/12, at 4 pm

CS 188: Artificial Intelligence

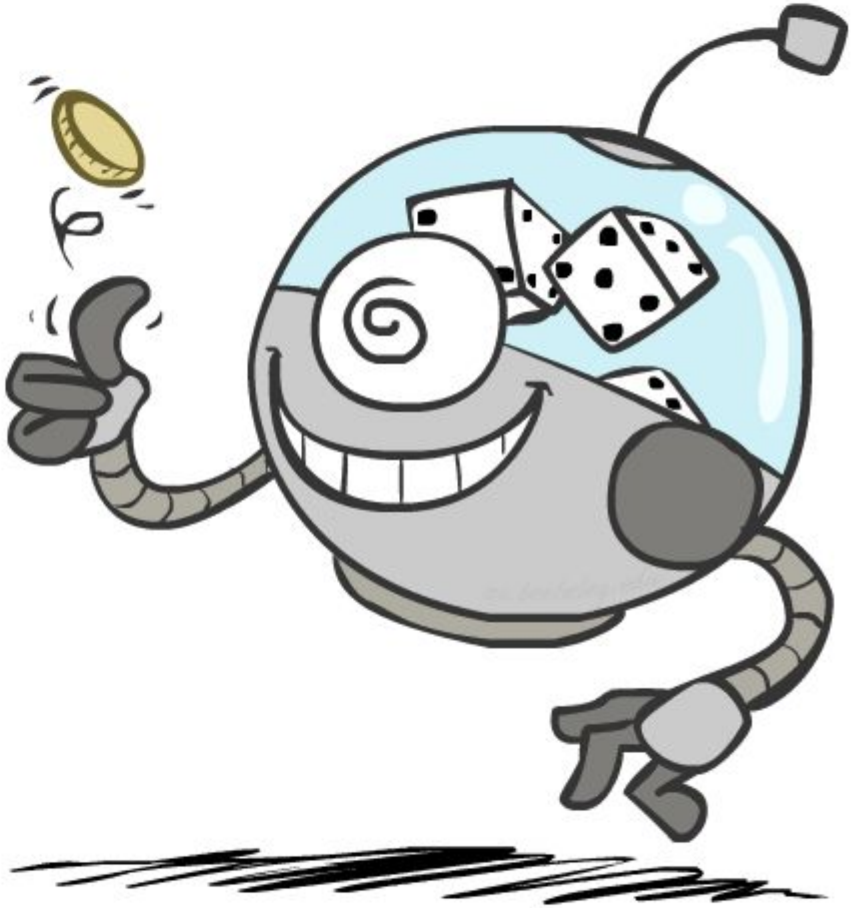
Probability

Instructors: Aditya Baradwaj and Brijen Thananjeyan --- University of California, Berkeley

Today

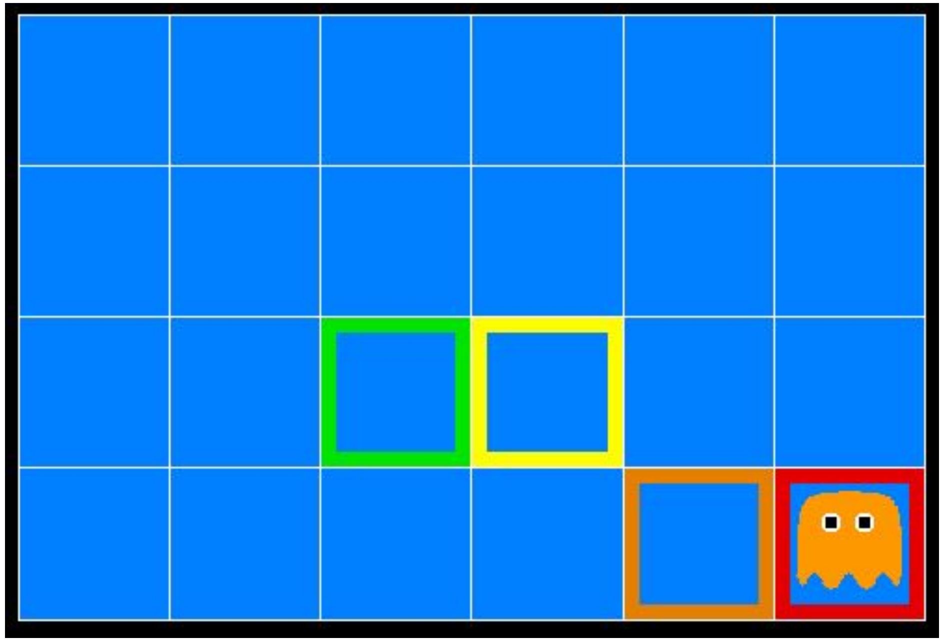
- Probability
  - Random Variables
  - Joint and Marginal Distributions
  - Conditional Distribution
  - Product Rule, Chain Rule, Bayes’ Rule
  - Inference
  - Independence

- You’ll need all this stuff A LOT for the next few weeks, so make sure you go over it now!



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 | Distance)

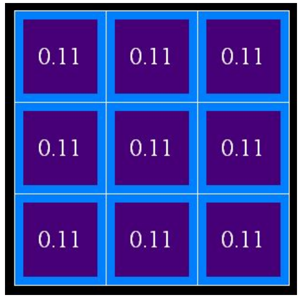
P(red | 3)	P(orange | 3)	P(yellow | 3)	P(green | 3)
0.05	0.15	0.5	0.3

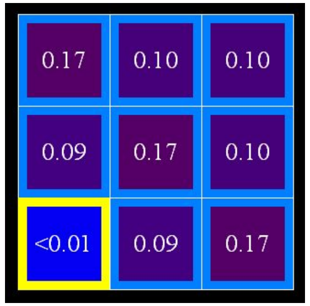
[Demo: Ghostbuster – no probability (L12D1)]

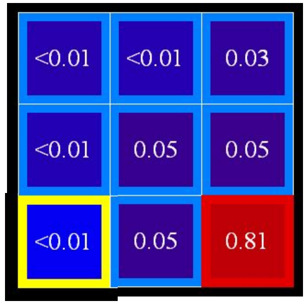
Video of Demo Ghostbuster

Uncertainty

- General situation:
  - **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
  - **Model**: Agent knows something about how the known variables relate to the unknown variables
- Probabilistic reasoning gives us a framework for managing *uncertain* beliefs and knowledge

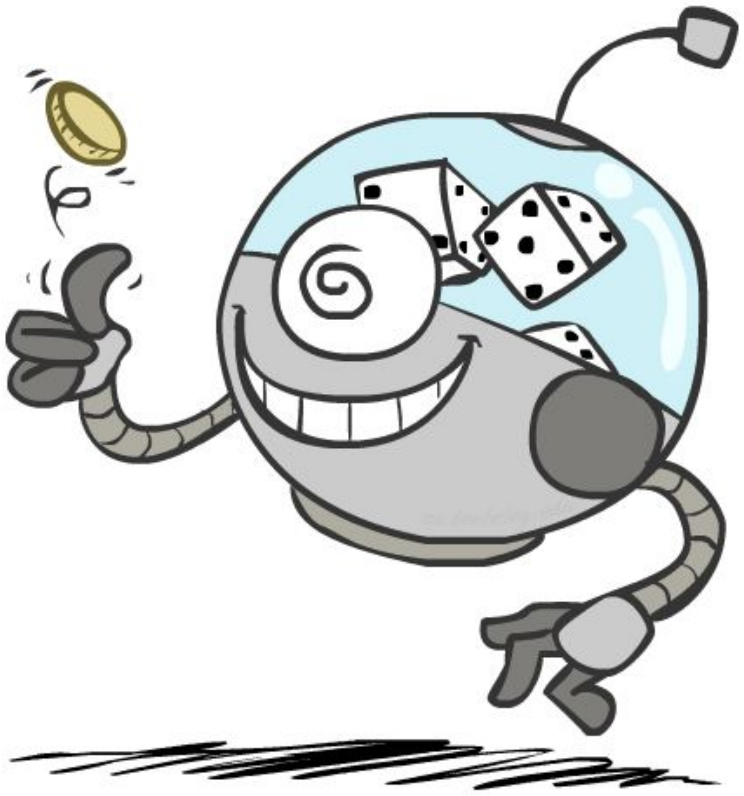






Random Variables

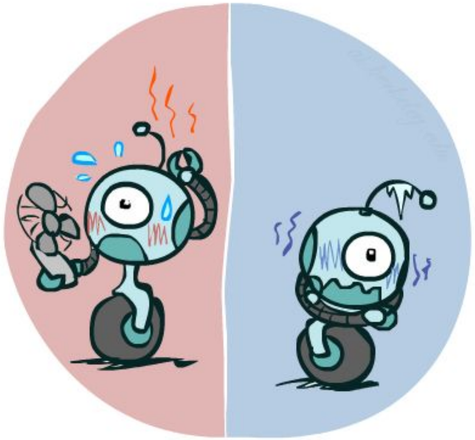
- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the ghost?
- (Technically, a random variable is a deterministic function from a possible world to some range of values.)
- 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

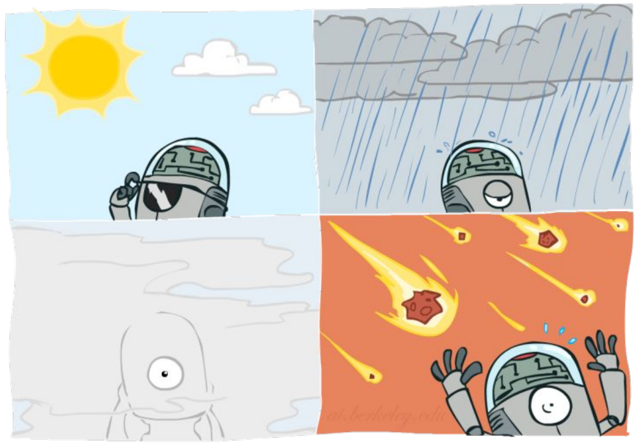
  - Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

  - Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

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

- A distribution for a discrete variable is a TABLE of probabilities of values

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

$$P(W = rain) = 0.1$$

- Must have: $\forall x \;\; P(X = x) \geq 0$ and $\sum_x P(X = x) = 1$

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

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*):

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

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

  - 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

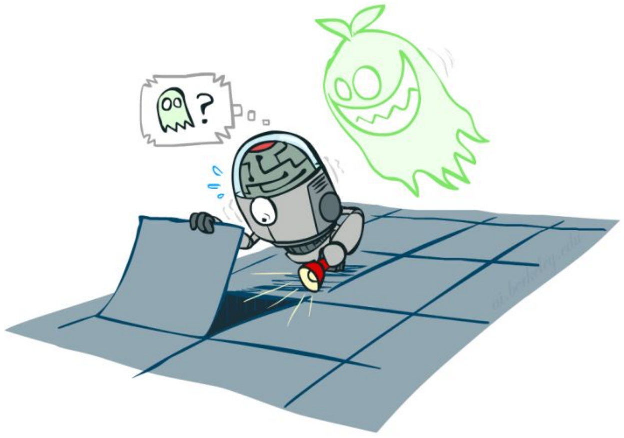
- Size of distribution if n variables with domain sizes d?
  - d^n. For all but the smallest distributions, impractical to write out!

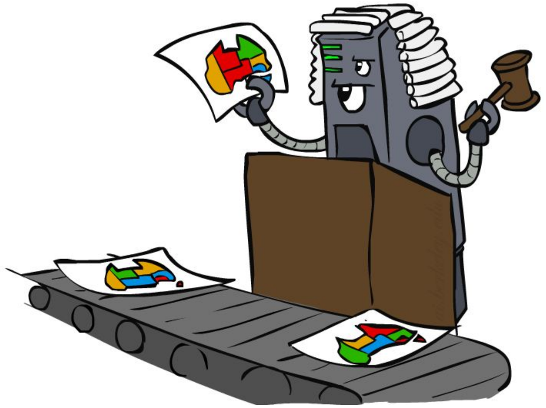
Probability Models

- A probability model is a joint distribution over a set of random variables
- Probability models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

Distribution over T,W

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





Events

- An *event* E is a set 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?
- 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) ?

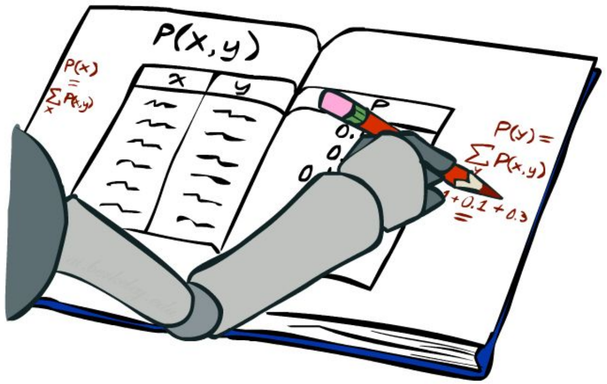
- P(+x) ?

- 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

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

$$P(t) = \sum_w P(t,w)$$

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

$P(T)$

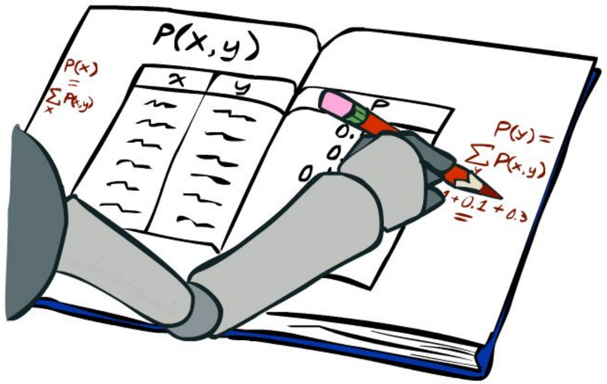
T	P
hot	0.5
cold	0.5

$P(W)$

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	
-x	

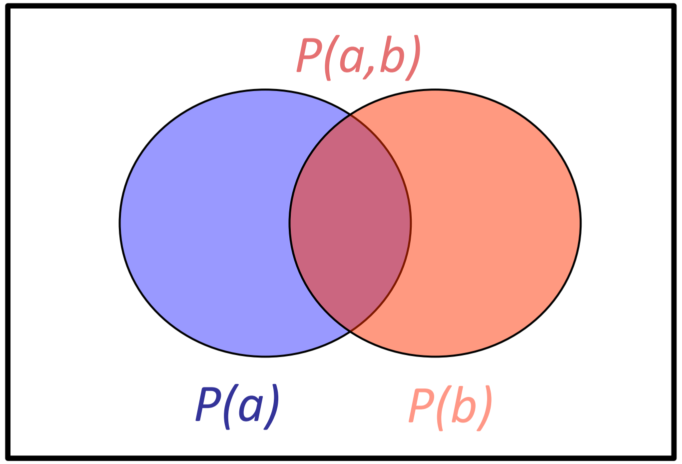
$P(Y)$

Y	P
+y	
-y	

Conditional Probabilities

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

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



$$P(T,W)$$

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

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

$$= P(W=s,T=c) + P(W=r,T=c)$$
$$= 0.2 + 0.3 \quad = 0.5$$

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) ?
- P(-x | +y) ?
- P(-y | +x) ?

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

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

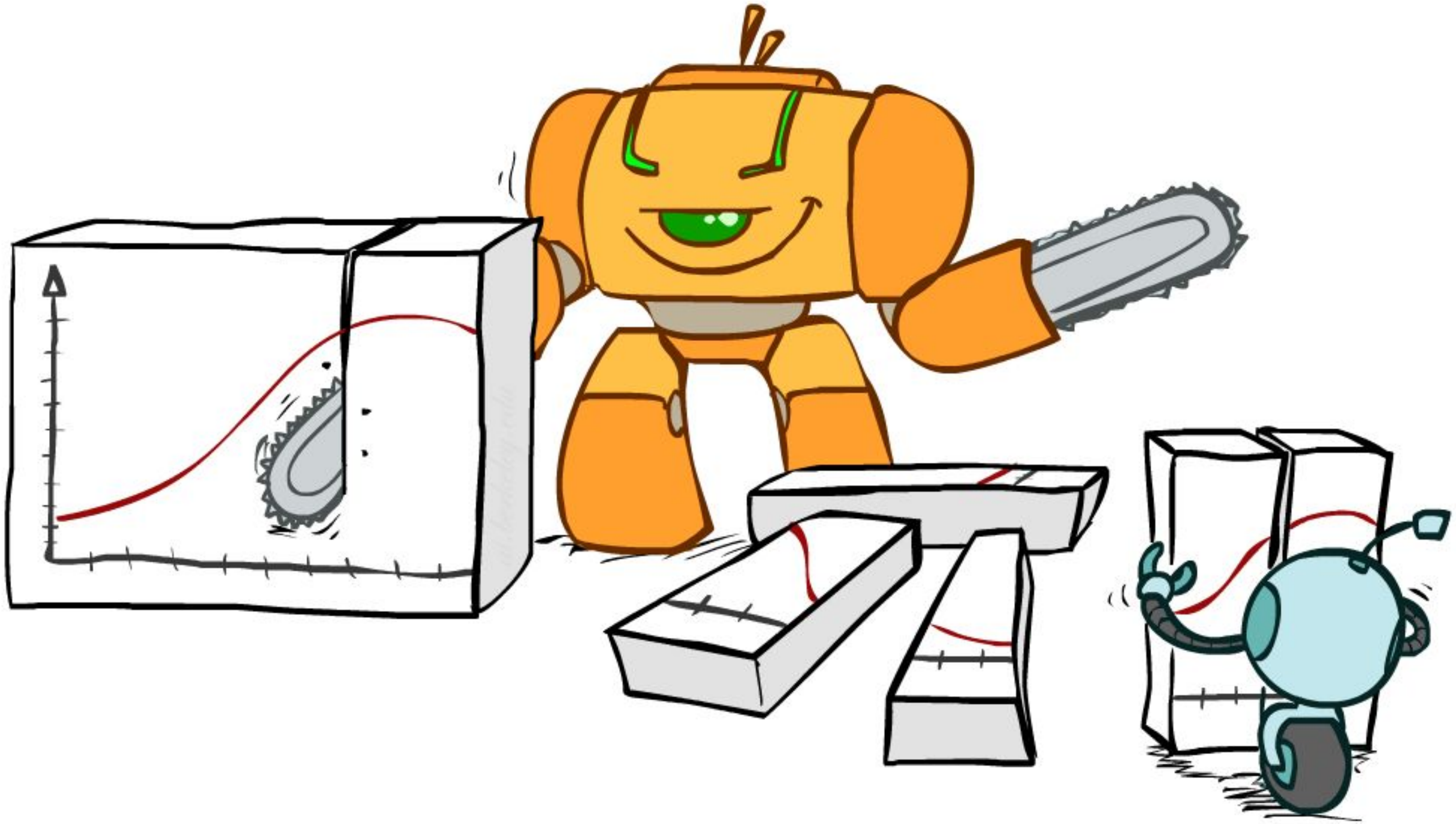
W	P
sun	0.4
rain	0.6

Joint Distribution

$$P(T, W)$$

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

Bayes Rule



Bayes' Rule

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

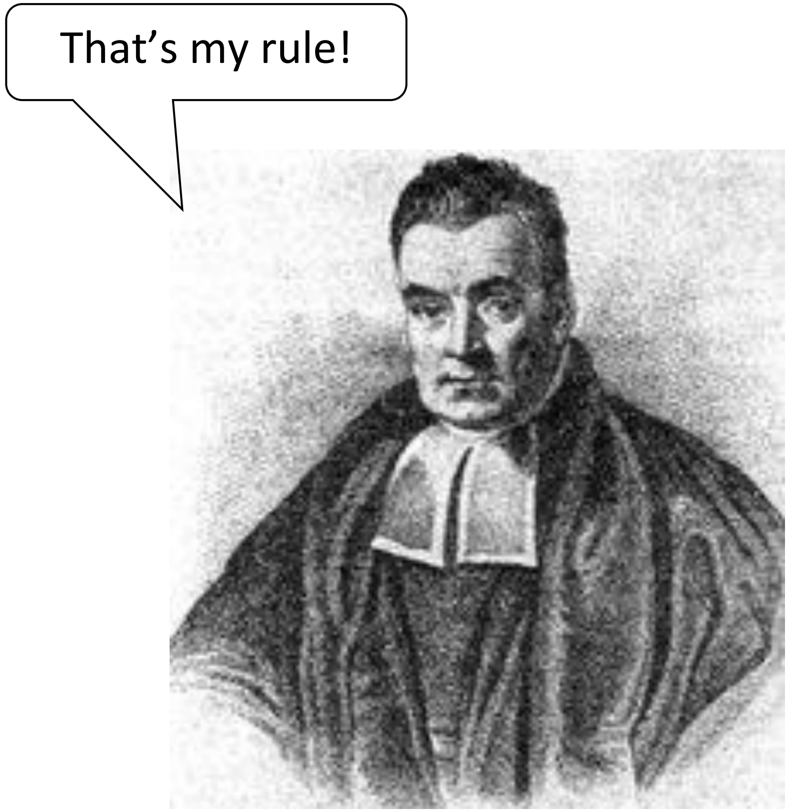
$$P(A,B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

- Dividing, we get:

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

- 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
  - Foundation of many systems we'll see later (e.g. ASR, MT)

- In the running for most important AI equation!

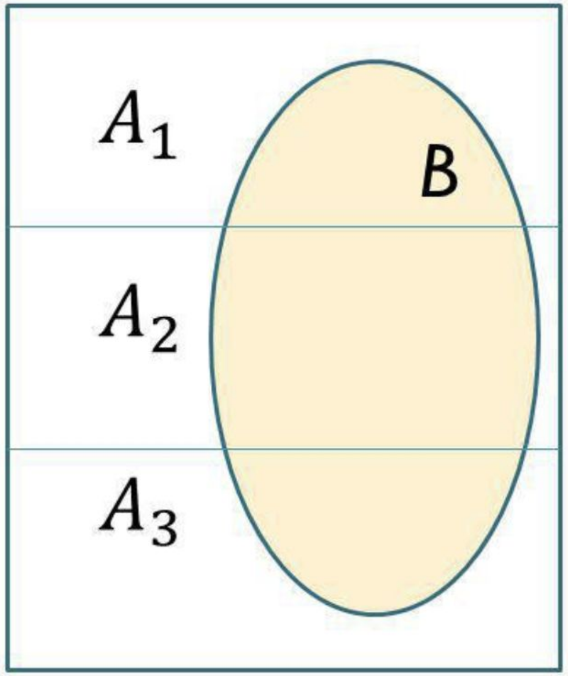


Law of total probability

- An event can be split up into it's intersection with disjoint events.

$$P(B) = P(B, A_1) + P(B, A_2) + \ldots + P(B, A_n)$$

  - Where A1, A2, ... An are mutually exclusive and exhaustive



Combining the two

- Here's what you get when you combine Bayes' Rule with the law of total probability:

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B)}$$ (Bayes Rule)

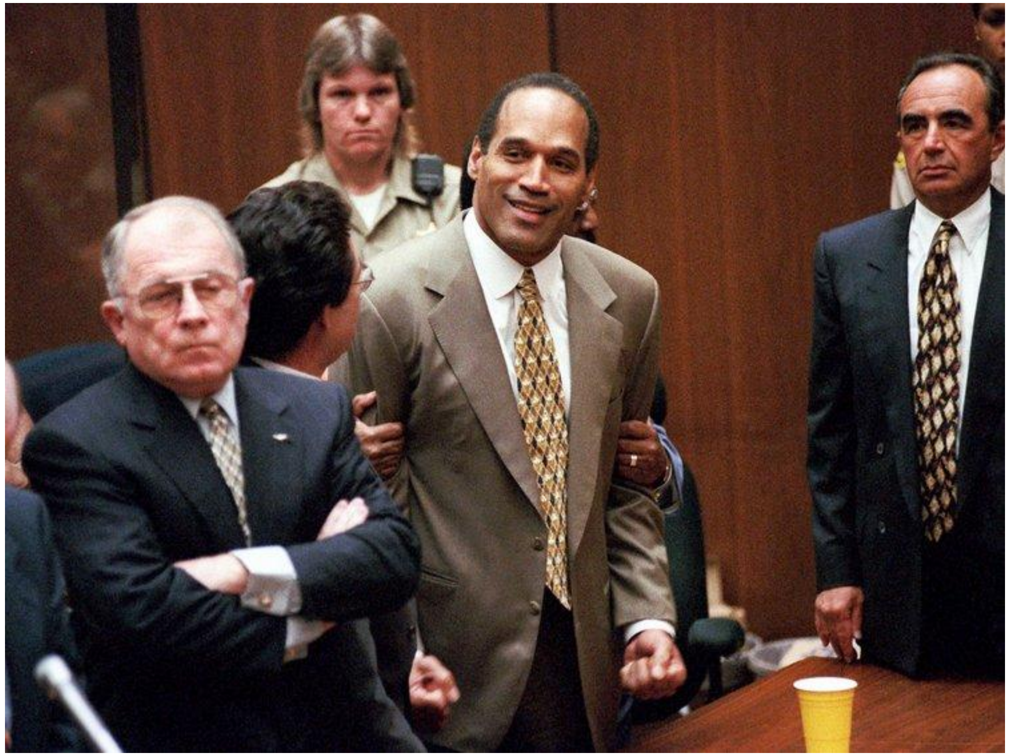
$$= \frac{P(B|A_1) \cdot P(A_1)}{P(B, A_1) + \ldots + P(B, A_n)}$$ (Law of total probability)

$$= \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1) \cdot P(A_1) + \ldots + P(B|A_n) \cdot P(A_n)}$$ (Bayes' Rule)

Case Study!

- OJ Simpson murder trial, 1995
  - “Trial of the Century”
  - OJ was suspected of murdering his wife and her friend.
  - Mountain of evidence against him

(DNA, bloody glove, history of abuse toward his wife).



Case Study

- Defense lawyer: Alan Dershowitz
- “Only one in a thousand abusive husbands eventually murder their wives.”
- Prosecution wasn’t able to convince the judge, and OJ was acquitted (he didn’t get charged)

Case Study

- Let's define the following events:
  - **M** – A wife is **M**urdered
  - **H** – A wife is murdered by her **H**usband
  - **A** – The husband has a history of **A**buse towards the wife

- Dershowitz' claim: "Only one in a thousand abusive husbands eventually murder their wives."
  - Translates to

$$P(H|A) = \frac{1}{1000}$$

- Dershowitz' claim: "Only one in a thousand abusive husbands eventually murder their wives."
  - Translates to $$P(H|A) = \frac{1}{1000}$$
- Does anyone see the problem here?
- But we don't care about P(H | A), we want P(H | A, M)
  - Why?
  - Since we know the wife has been murdered!

Case Study

- Let's see if we can do a better job than the prosecution!

- Given:
  - $P(M|H) = P(M|H,A) = 1$
  - $P(M|\bar{H})$: in 1994, 5000 women were murdered, 1500 by their husband. Assuming a population of 100 million women, we have:
    - $P(M|\bar{H}) = P(M|\bar{H},A) = \frac{3500}{100 \times 10^6} \approx \frac{1}{30{,}000}$
  - $P(H|A) = {}^{1}/_{1000}$
  - $P(\bar{H}|A) = {}^{999}/_{1000}$

- Use these to calculate:
  - $P(H|A)$

$$
\begin{aligned}
P(H|A,M) &= \frac{P(H,A,M)}{P(A,M)} && \text{(Bayes' Theorem)} \\
&= \frac{P(M|H,A)\cdot P(H|A)\cdot P(A)}{P(M|A)\cdot P(A)} && \text{(Bayes' Theorem)} \\
&= \frac{P(M|H,A)\cdot P(H|A)}{P(M|A)} && \text{(Cancellation)} \\
&= \frac{P(M|H,A)\cdot P(H|A)}{P(M,H|A)+P(M,\overline{H}|A)} && \text{(Law of total probability)} \\
&= \frac{P(M|H,A)\cdot P(H|A)}{P(M|H,A)\cdot P(H|A)+P(M|\overline{H},A)\cdot P(\overline{H}|A)} && \text{(Bayes' Theorem)} \\
&= \frac{1\cdot 1/1000}{1\cdot 1/1000+1/30000\cdot 999/1000} && \text{(Substituting given values)} \\
&= \frac{30,000}{30,999} = 0.97 = 97\%
\end{aligned}
$$

Case Study

- 97% probability that OJ murdered his wife!
- Quite different from 0.1%

- Maybe if the prosecution had realized this, things would have gone differently.

- Moral of the story: know your conditional probability!

Break!

- Stand up and stretch
- Talk to your neighbors

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

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

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

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

Quiz: Bayes' Rule

- Given:

$P(W)$

W	P
sun	0.8
rain	0.2

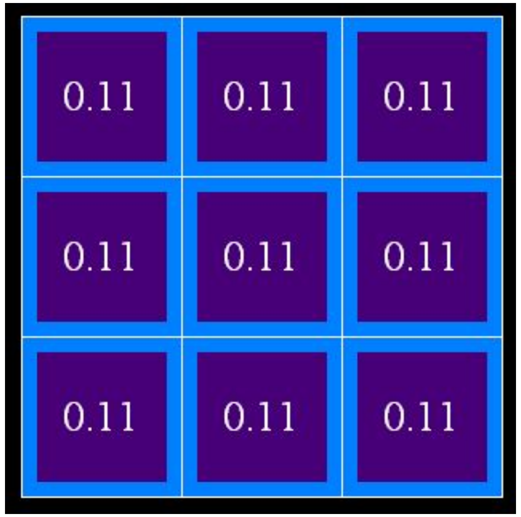
$P(D|W)$

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

- What is P(W | dry) ?

Ghostbusters, Revisited

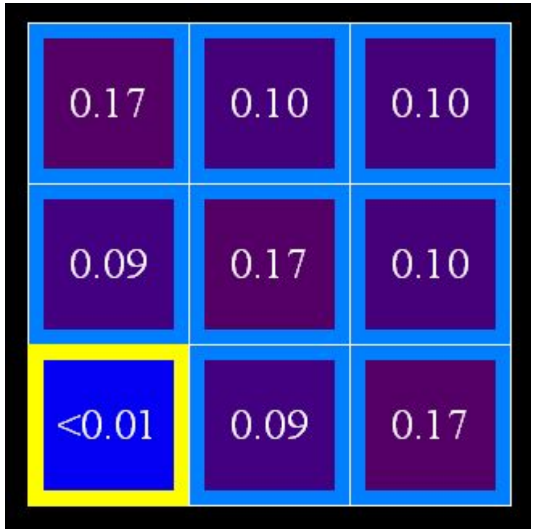
- Let's say we have two distributions:
  - Prior distribution over ghost location: P(G)
    - Let's say this is uniform
  - Sensor reading model: P(R | G)
    - Given: we know what our sensors do
    - R = reading color measured at (1,1)
    - E.g. P(R = yellow | G=(1,1)) = 0.1



- We can calculate the posterior distribution P(G|r) over ghost locations given a reading using Bayes' rule:

$$P(g|r) \propto P(r|g)P(g)$$

- What about two readings? What is $P(r_1, r_2 \mid g)$?



To Normalize

- (Dictionary) To bring or restore to a normal condition

All entries sum to ONE

- Procedure:
  - Step 1: Compute Z = sum over all entries
  - Step 2: Divide every entry by Z

- Example 1

W	P
sun	0.2
rain	0.3

Normalize $\xrightarrow{\quad}$ Z = 0.5

W	P
sun	0.4
rain	0.6

- Example 2

T	W	P
hot	sun	20
hot	rain	5
cold	sun	10
cold	rain	15

Normalize $\xrightarrow{\quad}$ Z = 50

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

Normalization Trick

- A trick to get a whole conditional distribution at once:
  - Select the joint probabilities matching the evidence
  - Normalize the selection (make it sum to one)

$P(T, W)$

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

Select

$P(T, r)$

T	R	P
hot	rain	0.1
cold	rain	0.3

Normalize

$P(T|r)$

T	P
hot	0.25
cold	0.75

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

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

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: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

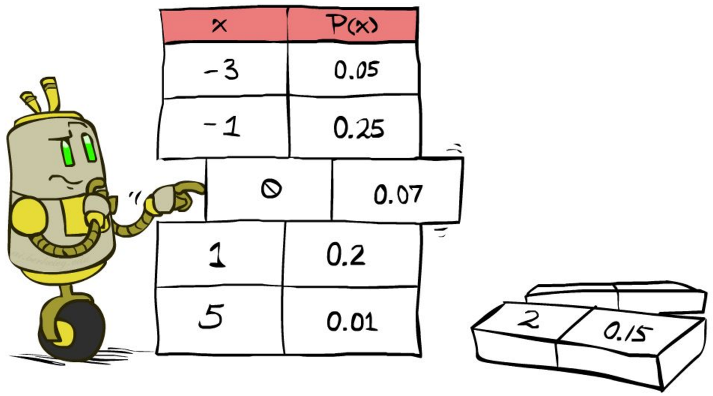
$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_1, X_2, \dots X_n$ *All variables*

- We want:

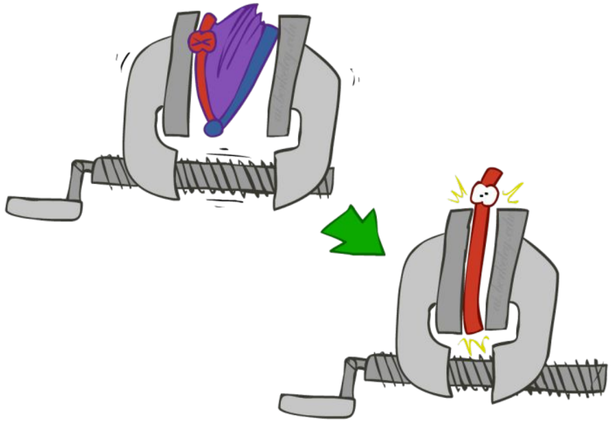
** Works fine with multiple query variables, too*

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

- 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

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

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

Inference by Enumeration

- P(W)?

- P(W | winter)?

- P(W | winter, hot)?

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

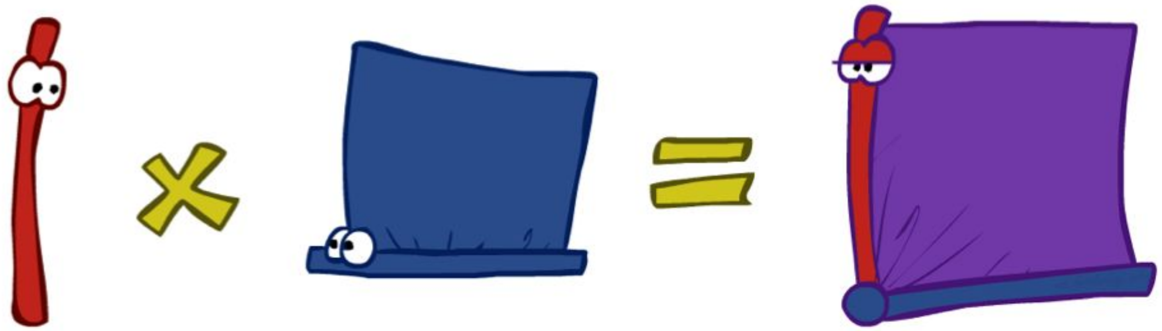
Inference by Enumeration

- 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

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



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	
dry	sun	
wet	rain	
dry	rain	

The Chain Rule

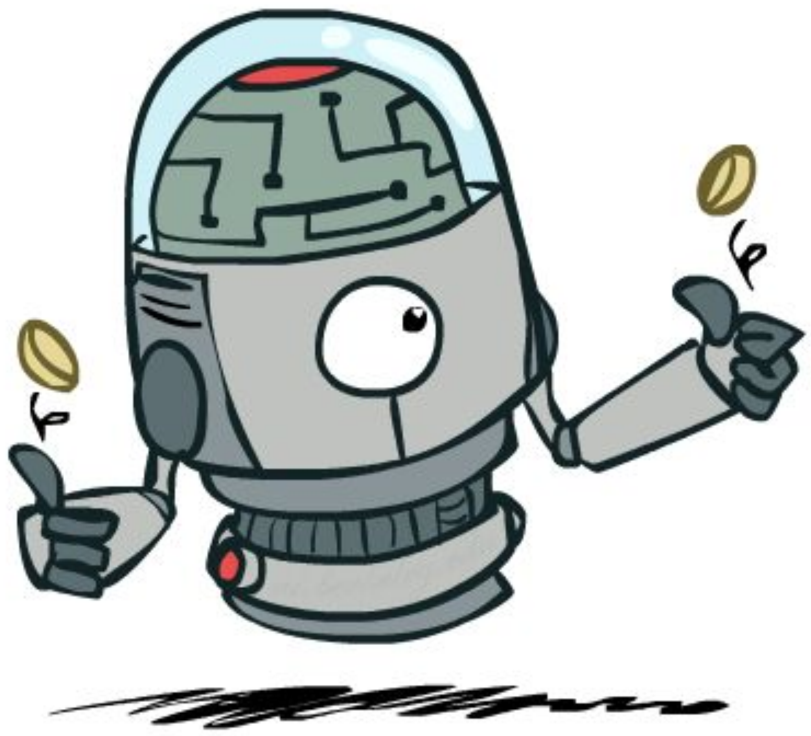
- More generally, can always write any joint distribution as an incremental product of conditional distributions

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

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

- Why is this always true?

Independence



Independence

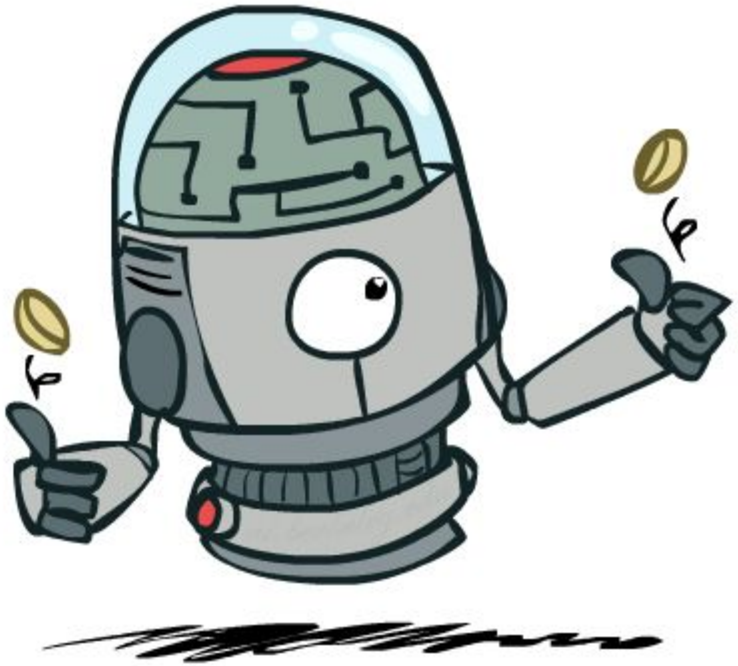
- Two variables are *independent* if:

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

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

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

- 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

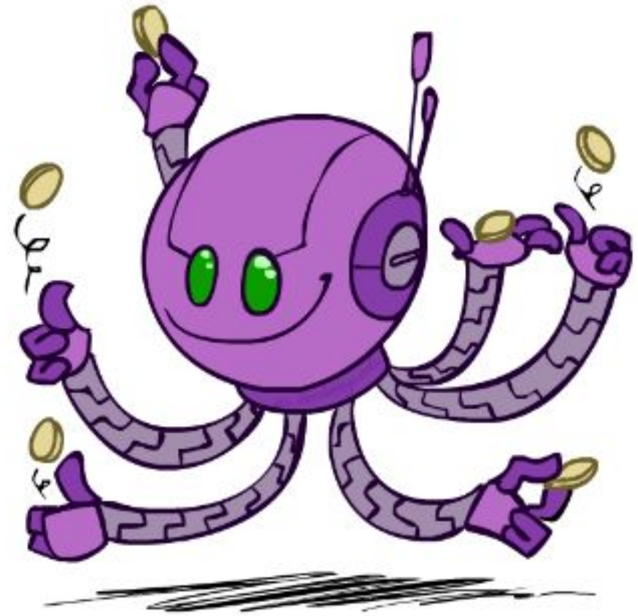
$P(X_2)$

H	0.5
T	0.5

$\ldots$

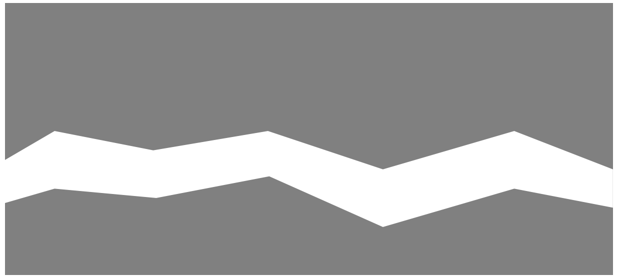
$P(X_n)$

H	0.5
T	0.5

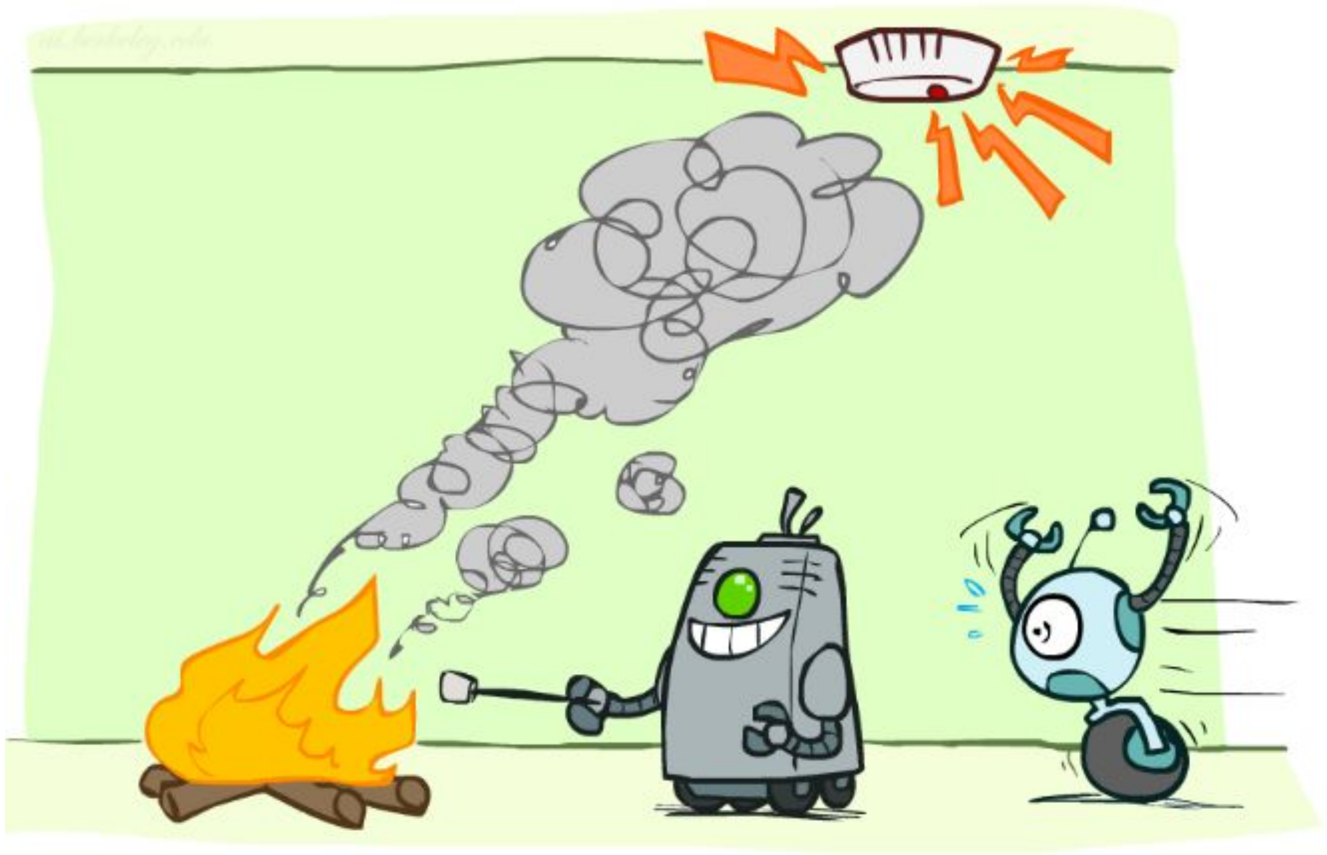


$P(X_1, X_2, \ldots X_n)$

2^n



Conditional Independence



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

  if and only if:

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

  or, equivalently, if and only if

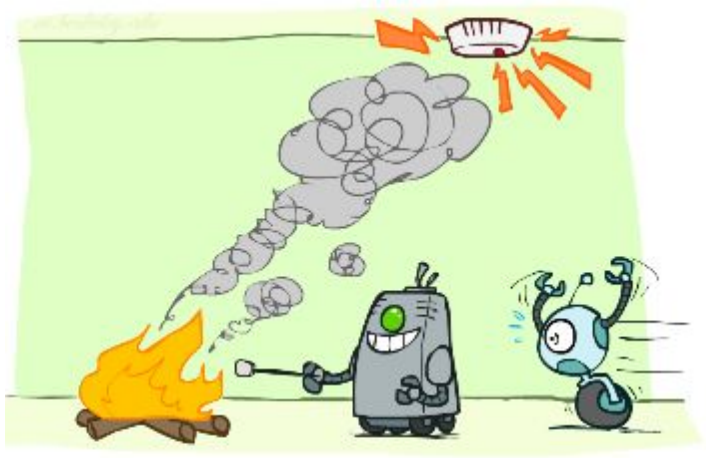
$$\forall x, y, z : P(x|z, y) = P(x|z)$$

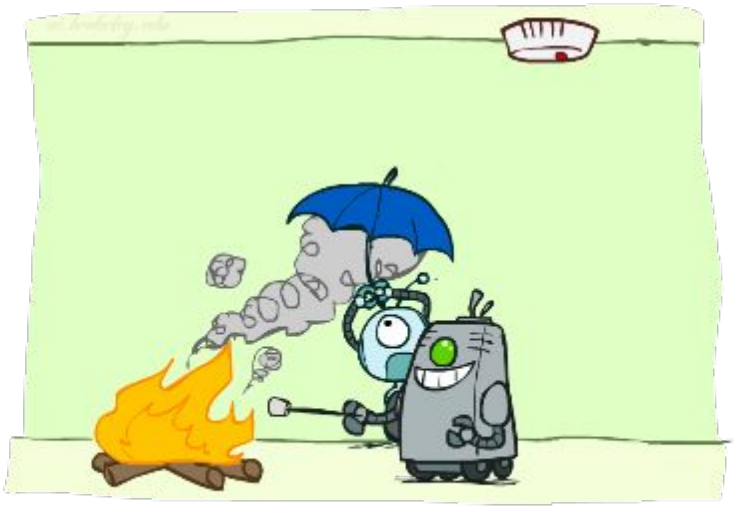
Conditional Independence

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

Conditional Independence

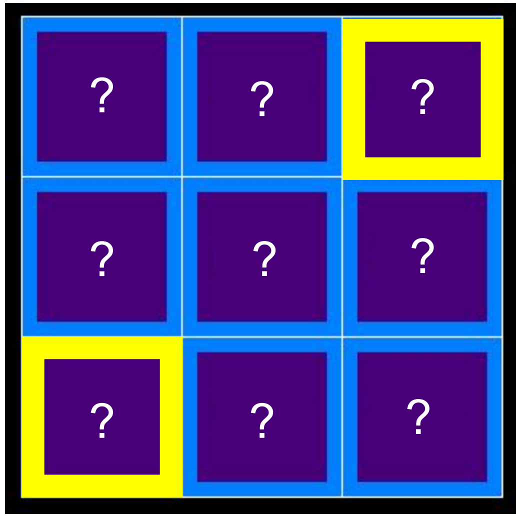
- What about this domain:
  - Fire
  - Smoke
  - Alarm

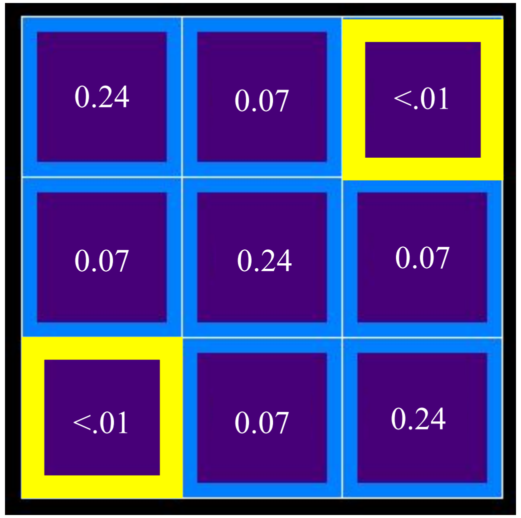




Ghostbusters, Revisited

- What about two readings? What is $P(r_1,r_2 \mid g)$?
- Readings are conditionally independent given the ghost location!
- $P(r_1,r_2 \mid g) = P(r_1 \mid g)\, P(r_2 \mid g)$
- Applying Bayes' rule in full:
- $P(g \mid r_1,r_2)$ α $P(r_1,r_2 \mid g)\, P(g)$
  $= P(g)\, P(r_1 \mid g)\, P(r_2 \mid g)$
- *Bayesian updating*





[Demo: Ghostbuster – with probability (L12D2)]

Video of Demo Ghostbusters with Probability

Summary

- Uncertainty is ubiquitous in the real world
- Probability theory is designed to handle uncertain information
  - "The theory of probabilities is just common sense reduced to calculus"

    Laplace, 1814
- A probability model assigns a probability to each possible world
  - Typically a Cartesian product of random variable assignments
- Any question can be answered by summing entries in the model
- Bayes' rule operates directly with conditional probabilities
- Independence and conditional independence simplify the model

Normalization Trick

$$P(T, W)$$

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

$$\begin{aligned} P(W = s | T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\ &= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.2}{0.2 + 0.3} = 0.4 \end{aligned}$$

$$\longrightarrow$$

$$\begin{aligned} P(W = r | T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\ &= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\ &= \frac{0.3}{0.2 + 0.3} = 0.6 \end{aligned}$$

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

W	P
sun	0.4
rain	0.6

Normalization Trick

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)}$$
$$= \frac{P(W=s,T=c)}{P(W=s,T=c)+P(W=r,T=c)}$$
$$= \frac{0.2}{0.2+0.3} = 0.4$$

$P(T,W)$

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

$$P(W=r|T=c) = \frac{P(W=r,T=c)}{P(T=c)}$$
$$= \frac{P(W=r,T=c)}{P(W=s,T=c)+P(W=r,T=c)}$$
$$= \frac{0.3}{0.2+0.3} = 0.6$$

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

→

NORMALIZE the selection (make it sum to one)

→