

# CSE 473: Artificial Intelligence

## Probability



Steve Tanimoto  
University of Washington

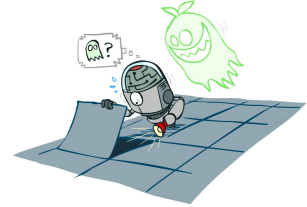
[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at <http://ai.berkeley.edu>.]

# Topics from 30,000'

▪ We're done with Part I Search and Planning!

▪ **Part II: Probabilistic Reasoning**

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



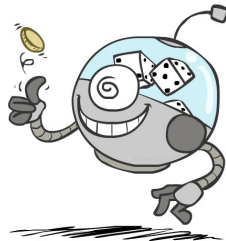
▪ **Part III: Machine Learning**

# Outline

▪ **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!



# Uncertainty

▪ **General situation:**

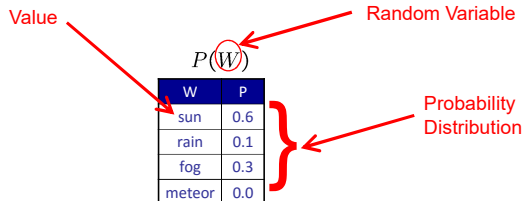
- **Observed variables (evidence):** Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Unobserved variables:** Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
  - **Model:** Agent knows something about how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

0.11	0.11	0.1
0.11	0.11	0.1
0.11	0.11	0.1

0.17	0.10	0.10
0.09	0.17	0.10
0.09	0.09	0.17

0.03	0.03	0.13
0.03	0.03	0.13
0.13	0.03	0.03

# What is....?



# Joint Distributions

▪ A *joint distribution* over a set of random variables:  $X_1, X_2, \dots, X_n$  specifies a probability for each assignment (or *outcome*):

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

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

- Must obey:  $P(x_1, x_2, \dots, x_n) \geq 0$
- $\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n) = 1$

$P(T, W)$

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

▪ Size of joint distribution if n variables with domain sizes d?

- For all but the smallest distributions, impractical to write out!

## Probabilistic Models

A *probabilistic model* is a joint distribution over a set of random variables

Distribution over T,W

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



Probabilistic models:

- (Random) variables with domains
- Joint distributions: say whether assignments (called "outcomes") are likely
- Normalized: sum to 1.0
- Ideally: only certain variables directly interact

Constraint satisfaction problems:

- Variables with domains
- Constraints: state whether assignments are possible
- Ideally: only certain variables directly interact

Constraint over T,W

T	W	P
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T



## Events

An *event* is a set E of outcomes

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

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?

$P(T, W)$

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

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

## Quiz: Events

•  $P(+x, +y)$  ?

•  $P(+x)$  ?

•  $P(-y \text{ OR } +x)$  ?

$P(X, Y)$

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

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

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

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

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

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

## 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	

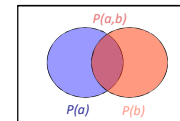
$P(Y)$

Y	P
+y	
-y	

## Conditional Probabilities

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

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



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

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

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

$$= 0.2 + 0.3 = 0.5$$

## 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

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

## 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

## Conditional Distributions - The Slow Way...

$P(T, W)$

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

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

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

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

$P(W|T = c)$

W	P
sun	0.4
rain	0.6

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

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

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

## Probabilistic Inference

- Probabilistic inference = *"compute a desired probability from other known probabilities (e.g. conditional from joint)"*
- We generally compute conditional probabilities
  - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
  - These represent the agent's *beliefs* given the evidence
- Probabilities change with new evidence:
  - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
  - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
  - 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:  $H_1 \dots H_r$
- We want:  $P(Q|e_1 \dots e_k)$
- \* Works fine with multiple query variables, too

Step 1: Select the entries consistent with the evidence

x	3x+5
-3	0.98
-1	0.25
0	0.07
1	0.2
5	0.05

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

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

Step 3: Normalize

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

$$Z = \sum_{q} P(Q, e_1 \dots e_k)$$

## Inference by Enumeration

- $P(W)$ ?
- $P(W \mid \text{winter})$ ?
- $P(W \mid \text{winter, hot})$ ?

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

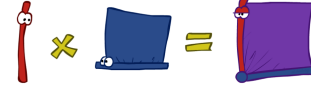
## Inference by Enumeration

- Computational 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

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



## The Product Rule

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

- Example:

$P(W)$		$P(D W)$			$P(D, W)$		
R	P	D	W	P	D	W	P
sun	0.8	wet	sun	0.1	wet	sun	
rain	0.2	dry	sun	0.9	dry	sun	
		wet	rain	0.7	wet	rain	
		dry	rain	0.3	dry	rain	

## The Chain Rule

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

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

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

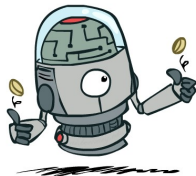
## Independence

- Two variables are *independent* in a joint distribution if:

$$P(X, Y) = P(X)P(Y)$$

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

$$X \perp\!\!\!\perp Y$$



- Says the joint distribution *factors* into a product of two simple ones
- Usually variables aren't independent!
- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?
- Independence is like something from CSPs: what?

## Example: Independence?

$P_1(T, W)$			$P(T)$		$P_2(T, W) = P(T)P(W)$		
T	W	P	T	P	T	W	P
hot	sun	0.4	hot	0.5	hot	sun	0.3
hot	rain	0.1	cold	0.5	hot	rain	0.2
cold	sun	0.2			cold	sun	0.3
cold	rain	0.3			cold	rain	0.2

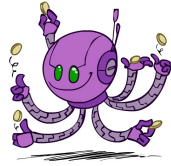
  

$P(W)$	
W	P
sun	0.6
rain	0.4

## Example: Independence

- N fair, independent coin flips:

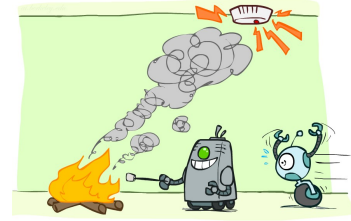
$P(X_1)$		$P(X_2)$		...	$P(X_n)$	
H	0.5	H	0.5		H	0.5
T	0.5	T	0.5		T	0.5



$$2^n \left\{ \begin{array}{c} P(X_1, X_2, \dots, X_n) \\ \text{[Graph showing a jagged line representing a probability distribution over a range]} \end{array} \right.$$

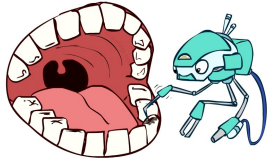


## Conditional Independence



## Conditional Independence

- $P(\text{Toothache}, \text{Cavity}, \text{Catch})$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - $P(+\text{catch} \mid +\text{toothache}, +\text{cavity}) = P(+\text{catch} \mid +\text{cavity})$
- The same independence holds if I don't have a cavity:
  - $P(+\text{catch} \mid +\text{toothache}, -\text{cavity}) = P(+\text{catch} \mid -\text{cavity})$
- Catch is *conditionally independent* of Toothache given Cavity:
  - $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$
- Equivalent statements:
  - $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
  - $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$
  - 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  $X \perp\!\!\!\perp Y \mid Z$   
if and only if:  

$$\forall x, y, z : P(x, y \mid z) = P(x \mid z) P(y \mid z)$$
 or, equivalently, if and only if  

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

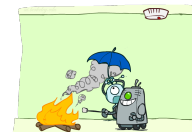
## Conditional Independence

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

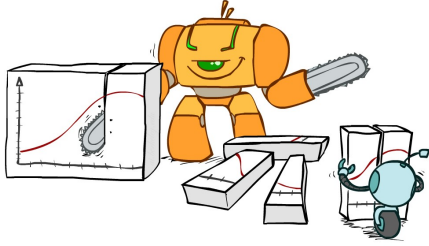


## Conditional Independence

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



## Bayes Rule



## Pacman – Sonar (P4)



[Demo: Pacman – Sonar – No Beliefs(L14D1)]

## Video of Demo Pacman – Sonar (no beliefs)



## Bayes' Rule

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

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

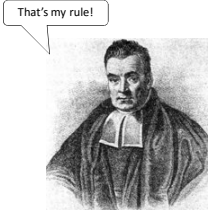
- Dividing, we get:

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

- 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!



## 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

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

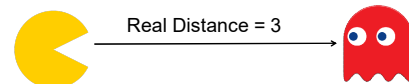
$$P(+m | +s) = \frac{P(+s | +m)P(+m)}{P(+s)} = \frac{P(+s | +m)P(+m)}{P(+s | +m)P(+m) + P(+s | -m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999} = 0.0079$$

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

## Ghostbusters Sensor Model

### Values of Pacman's Sonar Readings

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



36

## 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.  $P(R = \text{yellow} | G=(1,1)) = 0.1$

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

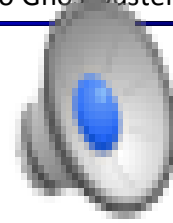
0.17	0.10	0.10
0.09	0.17	0.10
0.01	0.09	0.17

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

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

[Demo: Ghostbuster - with probability (L12D2)]

## Video of Demo Ghostbusters with Probability



## Probability Recap

- Conditional probability**  $P(x|y) = \frac{P(x,y)}{P(y)}$
- Product rule**  $P(x,y) = P(x|y)P(y)$
- Chain rule**  $P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$   
 $= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1})$
- Bayes rule**  $P(x|y) = \frac{P(y|x)}{P(y)}P(x)$
- X, Y independent if and only if:**  $\forall x, y : P(x, y) = P(x)P(y)$
- X and Y are conditionally independent given Z:**  $X \perp\!\!\!\perp Y | Z$   
 if and only if:  $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$