

CS 188: Artificial Intelligence

Fall 2011

Lecture 12: Probability

10/4/2011

Dan Klein – UC Berkeley

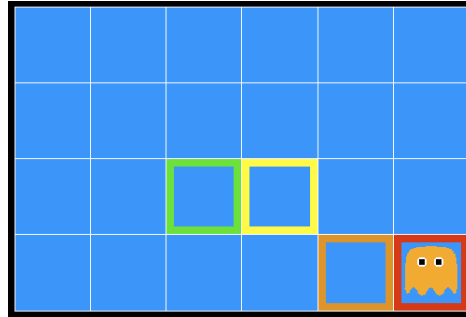
Today

- 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!

2

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 $P(\text{Color} \mid \text{Distance})$

$P(\text{red} \mid 3)$	$P(\text{orange} \mid 3)$	$P(\text{yellow} \mid 3)$	$P(\text{green} \mid 3)$
0.05	0.15	0.5	0.3

[Demo]

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	0.11
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

<0.01	<0.01	0.03
<0.01	0.05	0.05
<0.01	0.05	0.81

4

Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
 - R = Is it raining?
 - D = How long will it take to drive to work?
 - L = Where am I?
- We denote random variables with capital letters
- Like variables in a CSP, random variables have domains
 - R in {true, false} (sometimes write as {+r, -r})
 - D in $[0, \infty)$
 - L in possible locations, maybe $\{(0,0), (0,1), \dots\}$

5

Probability Distributions

- Unobserved random variables have distributions

$P(T)$		$P(W)$	
T	P	W	P
warm	0.5	sun	0.6
cold	0.5	rain	0.1
		fog	0.3
		meteor	0.0

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number

$$P(W = \text{rain}) = 0.1 \qquad P(\text{rain}) = 0.1$$

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

6

Joint Distributions

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

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

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

- Size of distribution if n variables with domain sizes d?
- 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

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

7

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
 - Assignments are called *outcomes*
 - Joint distributions: say whether assignments (outcomes) are likely
 - Normalized*: sum to 1.0
 - 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

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

8

Events

- An *event* is a set E of outcomes

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

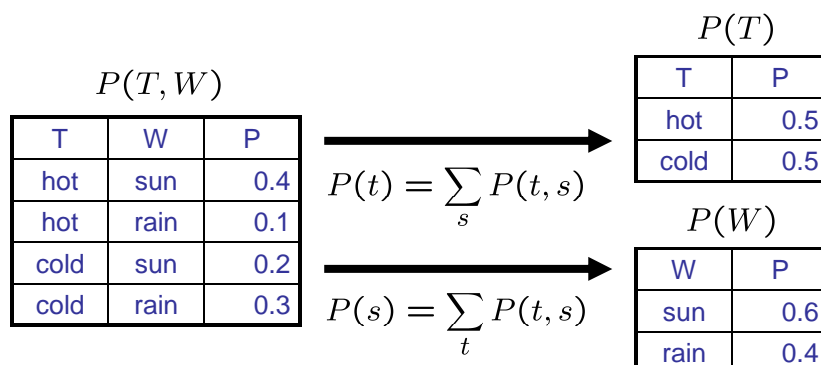
- From a joint distribution, we can calculate the probability of any event
 - Probability that it's hot AND sunny?
 - Probability that it's hot?
 - Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like $P(T=\text{hot})$

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

9

Marginal Distributions

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



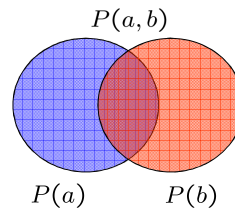
$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

10

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

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

11

Conditional Distributions

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

Conditional Distributions

Joint Distribution

$$P(W|T) \left\{ \begin{array}{l} P(W|T = \text{hot}) \\ P(W|T = \text{cold}) \end{array} \right.$$

W	P
sun	0.8
rain	0.2

W	P
sun	0.4
rain	0.6

$P(T, W)$

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

12

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 $P(\text{evidence})$! ($P(r)$, here)

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

13

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*

14

Inference by Enumeration

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

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

15

Inference by Enumeration

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

$\left. \begin{array}{l} X_1, X_2, \dots, X_n \\ \text{All variables} \end{array} \right\}$

- **We want:** $P(Q \mid e_1 \dots e_k)$
- First, select the entries consistent with the evidence
- Second, sum out H to get joint of Query and evidence:

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

- Finally, normalize the remaining entries to conditionalize

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

** Works fine with multiple query variables, too*

The Product Rule

- Sometimes have conditional distributions but want the joint

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

- 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	0.08
rain	0.2	dry	sun	0.9	dry	sun	0.72
		wet	rain	0.7	wet	rain	0.14
		dry	rain	0.3	dry	rain	0.06

The Chain Rule

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

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

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

- Why is this always true?

Bayes' Rule

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

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

That's my rule!

- 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!

20

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 is meningitis, s is stiff neck

$P(s m) = 0.8$	}	Example gives
$P(m) = 0.0001$		
$P(s) = 0.1$		

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

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

21

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

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

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

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

22

Independence

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

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

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

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

- 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?

23

Example: Independence?

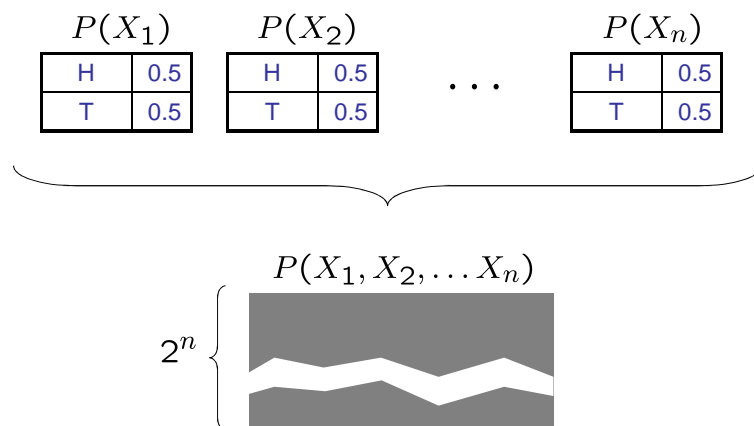
$P_1(T, W)$			$P(T)$		$P_2(T, W)$		
T	W	P	T	P	T	W	P
hot	sun	0.4	hot	0.5	hot	sun	0.3
hot	rain	0.1	cold	0.5	hot	rain	0.2
cold	sun	0.2			cold	sun	0.3
cold	rain	0.3			cold	rain	0.2

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

24

Example: Independence

- N fair, independent coin flips:



25