CS 188: Artificial Intelligence
Fall 2007

Lecture 14: Probability
10/11/2006

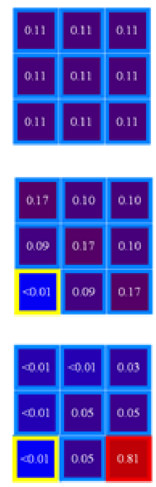
Dan Klein – UC Berkeley

Today

- Probability
  - Random Variables
  - Joint and Conditional Distributions
  - Bayes' Rule
  - Independence
- You'll need all this stuff for the next few weeks, so make sure you go over it!

Uncertainty

- General situation:
  - Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
  - 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



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 in a CSP, each random variable has a domain
  - R in {true, false}
  - D in $[0, \infty)$
  - L in possible locations

Probabilities

- We generally calculate conditional probabilities
  - P(on time | no reported accidents) = 0.90
- Probabilities change with new evidence:
  - P(on time | no reported accidents, 5 a.m.) = 0.95
  - P(on time | no reported accidents, 5 a.m., raining) = 0.80
  - i.e., observing evidence causes *beliefs to be updated*

Probabilistic Models

- CSPs:
  - Variables with domains
  - Constraints: map from assignments to true/false
  - Ideally: only certain variables directly interact
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: map from assignments (or outcomes) to positive numbers
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

A	B	P
warm	sun	T
warm	rain	F
cold	sun	F
cold	rain	T

A	B	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

Distributions on Random Vars

- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots X_n$ is a map from assignments (or *outcomes*, or *atomic events*) to reals:

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

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

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

$$0 \leq P(x_1, x_2, \ldots x_n) \leq 1$$

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

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

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

Examples

- An *event* is a set E of assignments (or 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 warm AND sunny?
- Probability that it's warm?
- Probability that it's warm OR sunny?

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

Marginalization

- Marginalization (or summing out) is *projecting* a joint distribution to a sub-distribution over subset of variables

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

$P(T, S)$

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

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

$P(T)$

T	P
warm	0.5
cold	0.5

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

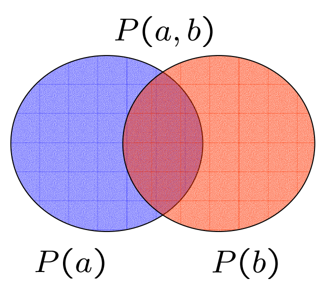
$P(S)$

S	P
sun	0.6
rain	0.4

Conditional Probabilities

- A conditional probability is the probability of an event given another event (usually evidence)

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



$P(T, S)$

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

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

Conditional Probabilities

- *Conditional* or *posterior probabilities:*
  - E.g., P(*cavity* | *toothache*) = 0.8
  - Given that *toothache* is all I know…
- Notation for conditional distributions:
  - P(*cavity* | *toothache*) = a single number
  - P(Cavity, Toothache) = 2x2 table summing to 1
  - P(Cavity | Toothache) = Two 2-element vectors, each summing to 1
- If we know more:
  - P(*cavity* | *toothache*, *catch*) = 0.9
  - P(*cavity* | *toothache*, *cavity*) = 1
- Note: the less specific belief remains *valid* after more evidence arrives, but is not always *useful*
- New evidence may be irrelevant, allowing simplification:
  - P(*cavity* | *toothache, traffic*) = P(*cavity* | *toothache*) = 0.8
- This kind of inference, guided by domain knowledge, is crucial

Conditioning

- Conditional distributions:

$$P(X_1|x_2) = \frac{P(X_1, x_2)}{P(x_2)}$$

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

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

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

Normalization Trick

- A trick to get the whole conditional distribution at once:
  - Get the joint probabilities for each value of the query variable
  - Renormalize the resulting vector

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

T	S	P
warm	sun	0.4
warm	rain	0.1
cold	sun	0.2
cold	rain	0.3

→ Select →

$P(T, r)$

T	P
warm	0.1
cold	0.3

→ Normalize →

$P(T|r)$

T	P
warm	0.25
cold	0.75

The Product Rule

- Sometimes joint P(X,Y) is easy to get
- Sometimes easier to get conditional P(X|Y)

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

- Example: P(sun, dry)?

$P(S)$

R	P
sun	0.8
rain	0.2

$P(D|S)$

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

⇔

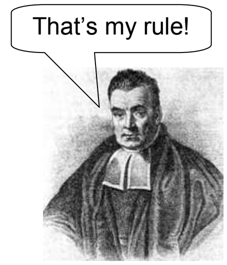
$P(D, S)$

D	S	P
wet	sun	0.08
dry	sun	0.72
wet	rain	0.14
dry	rain	0.06

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:

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

- Why is this at all helpful?
  - Lets us invert a conditional distribution
  - Often the one conditional is tricky but the other simple
  - Foundation of many systems we'll see later (e.g. ASR, MT)
- In the running for most important AI equation!

More Bayes' Rule

- 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

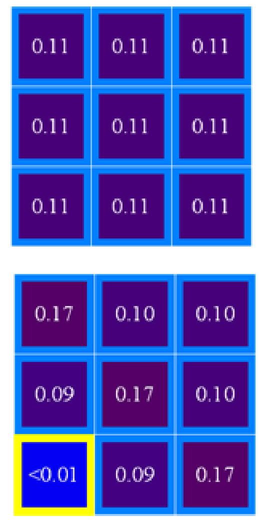
$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

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

Battleship

- Let's say we have two distributions:
  - Prior distribution over ship locations: P(L)
    - Say this is uniform
  - Sensor reading model: P(R | L)
    - Given by some known black box
    - E.g. P(R = yellow | L=(1,1)) = 0.1
    - For now, assume the reading is always for the lower left corner
- We can calculate the posterior distribution over ship locations using (conditionalized) Bayes' rule:

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



Inference by Enumeration

- P(sun)?
- P(sun | winter)?
- P(sun | winter, warm)?

S	T	R	P
summer	warm	sun	0.30
summer	warm	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	warm	sun	0.10
winter	warm	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- General case:
  - Evidence variables: $(E_1 \dots E_k) = (e_1 \dots e_k)$
  - Query variables: $Y_1 \dots Y_m$
  - Hidden variables: $H_1 \dots H_r$

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

- We want: $P(Y_1 \dots Y_m | e_1 \dots e_k)$

- First, select the entries consistent with the evidence
- Second, sum out H:

$$P(Y_1 \dots Y_m, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(\underbrace{Y_1 \dots Y_m, h_1 \dots h_r, e_1 \dots e_k}_{X_1, X_2, \dots X_n})$$

- 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