CS 188: Artificial Intelligence  
Spring 2009

Lecture 13: Probability  
03/03/09  
(square root day)

John DeNero – UC Berkeley  
Slides adapted from Dan Klein

Announcements

- **Project 3:**
  - Due tomorrow! Use up to 2 slip days.

- **Project 2:**
  - Should be graded by next Tuesday

- **Written Assignment 2:**
  - Longer than the last one
  - Due in lecture on Thursday 3/12
  - Slip days still do not apply to written assignments

Today

- **Probability**
  - Random Variables
  - Joint and Conditional Distributions
  - Inference, Bayes’ Rule
  - Independence

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

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

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

Sensors are noisy, but we know P(Color | Distance)

Uncertainty

- **General situation:**
  - **Evidence:** Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)
  - **Hidden 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

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 variable in a constraint satisfaction problem, each random variable has a domain
  - R in {true, false} (often write as \{r, ¬r\})
  - D in [0, x]
  - L in possible locations
### Probabilities
- We generally calculate conditional probabilities
  - \( P(\text{on time} | \text{no reported accidents}) = 0.90 \)
  - These represent the agent’s beliefs given the evidence
- Probabilities change with new evidence:
  - \( P(\text{on time} | \text{no accidents, 5 a.m.}) = 0.95 \)
  - Observing new evidence causes beliefs to be updated

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

### Joint Distributions
- A joint distribution over a set of random variables: \( X_1, X_2, \ldots, X_n \)
  - \( \sum_{x_1, x_2, \ldots, x_n} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = 1 \)
  - 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

### Events
- An event is a set \( E \) 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 \( P(T=h) \)

### Marginal Distributions
- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding
  - \( P(T, W) \)
  - \( P(T) = \sum_s P(t, s) \)
  - \( P(W) = \sum_t P(t, s) \)
  - \( P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) \)

### Conditional Distributions
- Conditional distributions are probability distributions over some variables given fixed values of others
  - \( P(W|T = \text{hot}) \)
  - \( P(W|T = \text{cold}) \)
Conditional Distributions

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

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

\[ P(T, W) \]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(W = r|T = c) = ??? \]

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)

\[ P(T, r) \]

Select

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Normalize

<table>
<thead>
<tr>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>0.25</td>
</tr>
<tr>
<td>cold</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- Why does this work? Because sum of selection is P(evidence)!

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

The Product Rule

- Sometimes have a joint distribution but want a conditional
  - Sometimes the reverse

\[ P(x|y) = \frac{P(x, y)}{P(y)} \leftrightarrow P(x, y) = P(x|y)P(y) \]

- Example:

\[ P(W) \]

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[ P(D|W) \]

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.08</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.14</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.08</td>
</tr>
</tbody>
</table>

\[ P(D, W) \]

Bayes’ Rule

- Two ways to factor a joint distribution over two variables:
  - 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 is meningitis, s is stiff neck
    - \[ P(s|m) = 0.8 \]
    - \[ P(m) = 0.0001 \]
    - \[ P(s) = 0.1 \]

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

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

Ghostbusters, Revisited

- Let’s say we have two distributions:
  - Prior distribution over ghost location: P(G)
    - Say this is uniform
  - Sensor reading model: P(R | G)
    - Given: we know what our sensors do
      - E.g. P(R = yellow | G = [1,1]) = 0.1
    - For now, assume the reading is always for the lower left corner

- We can calculate the posterior distribution over ghost locations given a reading using Bayes’ rule:

\[ P(\ell|r) \propto P(r|\ell)P(\ell) \]
Inference by Enumeration

- $P(\text{sun})$?

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

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

<table>
<thead>
<tr>
<th>S</th>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Inference by Enumeration

- General case:
  - Evidence variables: $(E_1 \ldots E_k) = (e_1 \ldots e_k)$
  - Query variables: $Y_1 \ldots Y_m$
  - Hidden variables: $H_1 \ldots H_n$
  - We want: $P(Y_1 \ldots Y_m \mid e_1 \ldots e_k)$

First, select the entries consistent with the evidence
Second, sum out $H$:
$$P(Y_1 \ldots Y_m, e_1 \ldots e_k) = \sum_{h_1 \ldots h_n} P(Y_1 \ldots Y_m, h_1 \ldots h_n, e_1 \ldots e_k)$$

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