CS 188: Artificial Intelligence
Fall 2009

Lecture 13: Probability
10/8/2009

Dan Klein – UC Berkeley

Announcements

- **Upcoming**
  - P3 Due 10/12
  - W2 Due 10/15
  - Midterm in evening of 10/22

- **Review sessions:**
  - Probability review: Friday 12-2pm in 306 Soda
  - Midterm review: on web page when confirmed
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 $P(\text{Color} | \text{Distance})$

| $P(\text{red} | 3)$ | $P(\text{orange} | 3)$ | $P(\text{yellow} | 3)$ | $P(\text{green} | 3)$ |
|-------------------|-------------------|-------------------|-------------------|
| 0.05              | 0.15              | 0.5               | 0.3               |
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 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), \ldots\}
Probability Distributions

- Unobserved random variables have distributions
  
  \[
  P(T) \quad P(W)
  \]
  
  \[
  \begin{array}{c|c}
  T & P \\
  \hline
  \text{warm} & 0.5 \\
  \text{cold} & 0.5 \\
  \end{array}
  \quad
  \begin{array}{c|c}
  W & P \\
  \hline
  \text{sun} & 0.6 \\
  \text{rain} & 0.1 \\
  \text{fog} & 0.3 \\
  \text{meteor} & 0.0 \\
  \end{array}
  \]

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number
  
  \[P(W = \text{rain}) = 0.1 \quad P(\text{rain}) = 0.1\]

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

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):
  
  \[P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)\]

  \[P(x_1, x_2, \ldots x_n)\]

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

- Must obey:
  
  \[P(x_1, x_2, \ldots x_n) \geq 0\]

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

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

<table>
<thead>
<tr>
<th>Distribution over T,W</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constraint over T,W</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>
Marginal Distributions

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

\[
P(T) = \sum_s P(t, s)
\]

\[
P(W) = \sum_t P(t, s)
\]

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

Conditional Probabilities

A simple relation between joint and conditional probabilities.

In fact, this is taken as the definition of a conditional probability:

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Conditional Distributions</th>
<th>Joint Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(W</td>
<td>T = \text{hot})$</td>
<td></td>
</tr>
<tr>
<td>$P(W</td>
<td>T = \text{cold})$</td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td>$P$</td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>rain</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

**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) \\
<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(T, r) \\
<table>
<thead>
<tr>
<th>T</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[
P(T|r) \\
<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? Sum of selection is $P$(evidence)! ($P(r)$, here)

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

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - \( P(\text{on time} \mid \text{no reported accidents}) = 0.90 \)
  - 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

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

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

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

<table>
<thead>
<tr>
<th></th>
<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>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
<td></td>
</tr>
</tbody>
</table>
Inference by Enumeration

- General case:
  - Evidence variables: \( E_1 \ldots E_k = e_1 \ldots e_k \)
  - Query* variable: \( Q \)
  - Hidden variables: \( H_1 \ldots H_r \)
  \[ X_1, X_2, \ldots, X_n \]

- We want: \( P(Q|e_1 \ldots e_k) \)

- First, select the entries consistent with the evidence

- Second, sum out \( H \) to get joint of Query and evidence:

\[
P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} P(Q, h_1 \ldots h_r, e_1 \ldots e_k)
\]

- 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

The Product Rule

- Sometimes have conditional distributions but want the joint

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

- Example:

| \( P(W) \) | \( P(D|W) \) | \( P(D, W) \) |
|---|---|---|
| R | P | D | W | P |
| sun | 0.8 | wet | sun | 0.1 |
| | | dry | sun | 0.9 |
| | | wet | rain | 0.7 |
| | | dry | rain | 0.3 |
| rain | 0.2 | wet | sun | 0.08 |
| | | dry | sun | 0.72 |
| | | wet | rain | 0.14 |
| | | dry | rain | 0.16 |

* Works fine with multiple query variables, too
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?

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

<table>
<thead>
<tr>
<th>m (meningitis)</th>
<th>s (stiff neck)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(s</td>
<td>m) = 0.8)</td>
</tr>
<tr>
<td>(P(m) = 0.0001)</td>
<td></td>
</tr>
</tbody>
</table>

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

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

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

\[
P(g|r) \propto P(r|g)P(g)
\]
Independence

- Two variables are independent in a joint distribution if:

\[ P(X, Y) = P(X)P(Y) \]
\[ \forall x, y \ P(x, y) = P(x)P(y) \]

- 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?

\[ P(T) \]
\[ P_1(T, W) \]
\[ P_2(T, W) \]
\[ P(W) \]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>warm</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>

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

<table>
<thead>
<tr>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.6</td>
</tr>
<tr>
<td>rain</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Example: Independence

- $N$ fair, independent coin flips:

<table>
<thead>
<tr>
<th></th>
<th>$P(X_1)$</th>
<th>$P(X_2)$</th>
<th>$P(X_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.5</td>
<td>H 0.5</td>
<td>H 0.5</td>
</tr>
<tr>
<td>T</td>
<td>0.5</td>
<td>T 0.5</td>
<td>T 0.5</td>
</tr>
</tbody>
</table>

$P(X_1, X_2, \ldots, X_n) = 2^n$