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
- 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 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

What is…?

- Random Variable
- Value
- Probability Distribution
- $P(X)$

Joint Distributions

- A joint distribution over a set of random variables $X_1, X_2, \ldots, X_n$ specifies a probability for each assignment (or outcome):
  $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
  $P(x_1, x_2, \ldots, x_n)$

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

- Size of joint distribution if $n$ variables with domain sizes $d_i$:
  - 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 are possible. "Outcomes"
  - 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

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

Events

- An event is a set E of outcomes
- From a joint distribution, we can calculate the probability of any event
- Typically, the events we care about are partial assignments, like P(T=hot)

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

Quiz: Events

- P(+x, +y)?
- P(+x)?
- P(-y OR +x)?

Marginal Distributions

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

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

Quiz: Marginal Distributions

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

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>+x</td>
<td>+y</td>
<td>0.2</td>
</tr>
<tr>
<td>+x</td>
<td>-y</td>
<td>0.3</td>
</tr>
<tr>
<td>-x</td>
<td>+y</td>
<td>0.4</td>
</tr>
<tr>
<td>-x</td>
<td>-y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Conditional Probabilities

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

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>W</td>
<td>P</td>
</tr>
<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 = s | T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4

Marginal: P(W = s | T = c) + P(W = r | T = c)

= P(W = s | T = c) + P(W = r | T = c)

= 0.2 + 0.3 = 0.5
Quiz: Conditional Probabilities

\[ P(X, Y) \]

\[ \begin{array}{ccc}
+X & +Y & 0.2 \\
+X & -Y & 0.3 \\
-X & +Y & 0.4 \\
-X & -Y & 0.1 \\
\end{array} \]

- \[ P(x \mid +y) ? \]
- \[ P(x \mid -y) ? \]
- \[ P(y \mid +x) ? \]

Conditional Distributions

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

\[ P(W|T = \text{hot}) \]

\[ P(W|T = \text{cold}) \]

Joint Distribution

\[ P(T, W) \]

- Normalization Trick

\[ P(T, W) \]

- Normalization Trick

\[ P(X, Y) \]

- Why does this work? Sum of selection is \( P(\text{evidence}) \) (\( P(T=\text{hot}) \), here):

\[ P(C|x) = \frac{P(x_1, x_2) \cdot P(C|x_1, x_2)}{P(C)} = \frac{\sum P(x_1, x_2)}{P(C)} \]
To Normalize

- Dictionary: “To bring or restore to normal condition
- Procedure:
  - Step 1: Compute $Z = \sum$ over all entries
  - Step 2: Divide every entry by $Z$

Example 1

<table>
<thead>
<tr>
<th>$W$</th>
<th>$P$</th>
<th>Normalize</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>rain</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>

Example 2

<table>
<thead>
<tr>
<th>$T$</th>
<th>$W$</th>
<th>$P$</th>
<th>Normalize</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

Probabilistic Inference

- Probabilistic inference = 
  compute a desired probability from other known probabilities (e.g., conditional from joint)
- We generally compute conditional probabilities
  - H(t) time | no reported accidents | 0.4
  - These represent the agent’s beliefs given the evidence
- Probabilities change with new evidence:
  - H(t) time | no accidents | 5 a.m.) | 0.5
  - H(t) time | no accidents | 5 a.m., raining | 0.5

Inference by Enumeration

- General case:
  - Evidence variables
  - Query variable
  - Hidden variables

We want:

\[
P(Q | c_1 \ldots c_k) = \frac{1}{Z} \prod_{i=1}^{k} P(Q | c_i) \prod_{i=1}^{n} P(Q_i | c_1 \ldots c_k, \text{all variables})
\]

Inference by Enumeration

<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>

<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>

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:

| \( P(W) \) | \( P(D|W) \) | \( P(D, W) \) |
|---|---|---|
| wet | sun | 0.1 |
| dry | sun | 0.9 |
| wet | rain | 0.7 |
| dry | rain | 0.3 |

The Chain Rule

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

\[ P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2) \]

\[ P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i|x_1, x_2, \ldots, x_{i-1}) \]

Independence

- Two variables are independent in a joint distribution if:

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

- Independence can be a simplifying assumption

- Independence is like something from CSPs: what?

Example: Independence?

| \( P(T) \) | \( P(W|T) \) | \( P(T, W) = P(T)P(W) \) |
|---|---|---|
| hot | sun | 0.08 |
| cold | sun | 0.72 |
| hot | rain | 0.14 |
| cold | rain | 0.06 |

Example: Independence

- \( N \) fair, independent coin flips:

\[ P(X_1) \quad P(X_2) \quad \ldots \quad P(X_n) \]

\[ 2^N \quad P(X_1, X_2, \ldots, X_n) \]

Conditional Independence
Conditional Independence

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  \[ P(+\text{catch} | +\text{toothache}, +\text{cavity}) = P(+\text{catch} | +\text{cavity}) \]
- The same independence holds if I don't have a cavity:
  \[ P(+\text{catch} | +\text{toothache}, -\text{cavity}) = P(+\text{catch} | -\text{cavity}) \]
- Catch is conditionally independent of Toothache given Cavity:
  \[ P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity}) \]
- Equivalent statements:
  - \[ P(\text{Toothache} | \text{Catch}, \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) \]
  - \[ P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity})P(\text{Catch} | \text{Cavity}) \]
  - One can be derived from the other easily.

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

\[ X \perp Y | Z \]

if and only if:
\[ \forall x, y, z : P(x, y | z) = P(x | z)P(y | z); \]
or, equivalently, if and only if
\[ \forall x, y, z : P(x | y, z) = P(x | z) \]

What about this domain:
- Traffic
- Umbrellas
- Raining

What about this domain:
- Fire
- Smoke
- Alarm

Bayes Rule

Pacman – Sonar (P4)

[Source: Pacman – Sonar – NicolerG2.1483]
**Ghostbusters Sensor Model**

Values of Pacman’s Sonar Readings

<table>
<thead>
<tr>
<th>Value</th>
<th>Sensor Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Red</td>
</tr>
<tr>
<td>2</td>
<td>Orange</td>
</tr>
<tr>
<td>3</td>
<td>Yellow</td>
</tr>
<tr>
<td>4</td>
<td>Green</td>
</tr>
</tbody>
</table>

Real Distance = 3

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

- Why is this at all helpful?
  
  - Let’s 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:
  
  \[ P(y|\neg e|x) = \frac{P(e|\neg y|x)P(y|x)}{P(e|x)} \]

- Example:
  
  - M: meningi\$s, S: stiff neck
  - \( \frac{P(S|\neg M)}{P(S|M)} = 0.0001 \)
  - \( P(S|\neg M) = 0.04 \)
  - \( P(S|M) = 0.040 \)

- Let’s build one conditional from its reverse

- Example:
  
  \[ P(S|\neg M) = \frac{P(S,M)}{P(S)} = \frac{P(S,M)}{P(S|M) + P(S,\neg M)} \]

- 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
  
  - 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:

  \[ P(G|R) \propto P(R|G)P(G) \]

**Ghostbusters, Revisited**

- Real Distance = 3

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

**Video of Demo Ghostbusters with Probability**
Probability Recap

- Conditional probability
  \[ P(x|y) = \frac{P(x, y)}{P(y)} \]
- Product rule
  \[ P(x, y) = P(y|x) P(y) \]
- Chain rule
  \[ P(x_1, x_2, \ldots, x_n) = P(x_1) P(x_2|x_1) P(x_3|x_2, x_3) \ldots = \prod_i P(x_i|x_{i-1}, \ldots, x_0) \]
- Bayes rule
  \[ P(x|y) = \frac{P(y|x) P(x)}{P(y)} \]

- \( X, Y \) independent if and only if: \( \forall x, y : P(x, y) = P(x) P(y) \)
- \( X \) and \( Y \) are conditionally independent given \( Z \): \( X \perp Y | Z \)
  if and only if: \( \forall x, y, z : P(x, y|z) = P(x|z) P(y|z) \)