CSE 473: Artificial Intelligence
Probability

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
  - Inference
  - 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 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

What is….?

Random Variable

Value

Probability Distribution

<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>0.1</td>
</tr>
<tr>
<td>fog</td>
<td>0.3</td>
</tr>
<tr>
<td>meteo</td>
<td>0.0</td>
</tr>
</tbody>
</table>

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):

\[
P(\{x_1, x_2, \ldots, x_n\}) = \frac{P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{P(x_1, x_2, \ldots, x_n)}
\]

- 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 \( d \)?
  - For all but the smallest distributions, impractical to write out!

<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>
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 (called "outcomes") are likely
  - 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

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 $P(T=\text{hot})$

Quiz: Events

- $P(\text{+x, +y})$?
- $P(\text{+x})$?
- $P(\text{-y OR +x})$?

Marginal Distributions

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

Quiz: Marginal Distributions

- $P(X, Y)$
  \[ P(X) = \sum_{Y} P(X, Y) \]
  \[ P(Y) = \sum_{X} P(X, Y) \]

Conditional Probabilities

- A simple relation between joint and marginal probabilities
  - In fact, this is taken as the definition of a conditional probability
  \[ P(a | b) = \frac{P(a, b)}{P(b)} \]
  \[ P(T, W) \]
  \[ P(T|W) = \frac{P(T, W)}{P(W)} \]
  \[ P(W|T) = \frac{P(T, W)}{P(T)} \]
  \[ P(W) = P(W|T = c) + P(W|T = \bar{c}) = 0.2 + 0.3 = 0.5 \]
**Quiz: Conditional Probabilities**

P(X, Y)

<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>

• P(+x | +y) ?
• P(-x | +y) ?
• P(-y | +x) ?

**Conditional Distributions**

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

**Joint Distribution**

P(W) 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>

**Probabilistic Inference**

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

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent’s beliefs given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes beliefs to be updated

**Inference by Enumeration**

- General case:
  - Evidence variables: E₁, E₂, ..., Eₜ
  - Query variable: Q
  - Hidden variables: H₁, H₂, ...

- We want: P(Q | e₁, ..., eₜ)

- Step 1: select the entries consistent with the evidence

- Step 2: sum out H to get joint of Query and evidence

- Step 3: Normalize

P(Q | e₁, ..., eₜ) = \frac{1}{Z} \prod Y_{i=1}^{t} P(Q | e_i, H_i) P(H_i)

Z = \sum_{Y} \prod_{i=1}^{t} P(Q | e_i, H_i) P(H_i)
Inference by Enumeration

- Computational 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

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

The Product Rule

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

- Example:

| $P(D|W)$ | $P(D,W)$ |
|----------|----------|
| sun      | wet sun 0.1 | wet sun 0.08 |
|         | dry sun 0.9 | dry sun 0.72 |
| rain     | wet rain 0.3 | wet rain 0.14 |
|          | dry rain 0.7 | dry rain 0.06 |

The Chain Rule

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

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

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

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 least “close” to independent
  - What could we assume for {Weather, Traffic, Cavity}?
- Independence is like something from CSPs: what?

Example: Independence?

<table>
<thead>
<tr>
<th>$P(T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
</tr>
<tr>
<td>cold</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P(T, W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot sun</td>
</tr>
<tr>
<td>hot rain</td>
</tr>
<tr>
<td>cold sun</td>
</tr>
<tr>
<td>cold rain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P(W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
</tr>
<tr>
<td>rain</td>
</tr>
</tbody>
</table>
Example: Independence

- N fair, independent coin flips:

<table>
<thead>
<tr>
<th></th>
<th>P(X₁)</th>
<th>P(X₂)</th>
<th>...</th>
<th>P(Xₙ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.5</td>
<td>H</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>T</td>
<td>0.5</td>
<td>T</td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

\[ P(X₁, X₂, \ldots, Xₙ) \]

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:
  \[ P(+\text{catch} | \text{toothache}, \text{cavity}) = P(+\text{catch} | \text{cavity}) \]

- Unconditional (absolute) independence very rare (why?)

- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

- If \( X \) is conditionally independent of \( Y \) given \( Z \):
  \[ X \perp Y | Z \]

  \[ P(x, y | z) = P(x | z)P(y | z) \]
  \[ \text{or, equivalently, if and only if} \]
  \[ P(x | y, z) = P(x | z) \]

What about this domain:

- Traffic
- Umbrella
- Raining

What about this domain:

- Fire
- Smoke
- Alarm
Bayes Rule

Pacman – Sonar (P4)

Video of Demo Pacman – Sonar (no beliefs)

Inference with Bayes’ Rule

Ghostbusters Sensor Model

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

• 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!

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: meningitis, S: stiff neck

\[
\begin{align*}
P(+m) &= 0.0001 \\
P(+s|m) &= 0.8 \\
P(+s|m') &= 0.01 \\
\end{align*}
\]

\[
\begin{align*}
P(+m; +s) &= P(+m)P(+s|m) \quad \text{Example} \\
&= 0.0001 \times 0.8 \\
&= 0.00008 \\
\end{align*}
\]

- Real distance = 3

Note: posterior probability of meningitis still very small

Note: you should still get stiff necks checked out! Why?

Values of Pacman’s Sonar Readings

<table>
<thead>
<tr>
<th>Value</th>
<th>P(red)</th>
<th>P(orange)</th>
<th>P(yellow)</th>
<th>P(green)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.15</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 3     | 0.8 \times 0.0003 + 0.3 \times 0.999  

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: \( P(G) \)
    - Let’s say this is uniform
  - Sensor reading model: \( P(R \mid G) \)
    - Given: we know what our sensors do
    - \( R \) = reading color measured at (1,1)
    - E.g. \( P(R = \text{yellow} \mid G = (1,1)) = 0.1 \)

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

Probability Recap

- Conditional probability
  \[
P(x \mid y) = \frac{P(x, y)}{P(y)}
  \]

- Product rule
  \[
P(x, y) = P(y \mid x)P(x)
  \]

- Chain rule
  \[
P(x_1, x_2, \ldots, x_n) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_2, x_3)\ldots
  = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1)
  \]

- Bayes rule
  \[
P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)}
  \]

- \( X, Y \) independent if and only if: \( \forall x, y : P(x, y) = P(x)P(y) \)

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