Announcements

- Assignments
  - P3 due tonight
  - W4 going out tonight

- Midterm
  - 3/18, 6-9pm, 0010 Evans
    - No lecture on 3/18
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

- $P(\text{sun}) = \sum_{S,T} P(S,T, \text{sun}) = 0.7 + 0.1 + 0.1 + 0.15 = 0.65$
- $P(\text{sun} \mid \text{winter}) = \frac{\sum_{S,T} P(S,T, \text{sun})}{\sum_{S,T} P(S,T, \text{winter})} = \frac{0.15}{0.1 + 0.05 + 0.15 + 0.2} = \frac{0.15}{0.5} = 0.30$
- $P(\text{sun} \mid \text{winter, warm}) = \frac{P(\text{sun, winter, warm})}{P(\text{winter, warm})} = \frac{0.1}{0.1 + 0.2} = \frac{0.1}{0.3} = \frac{1}{3}$
**Inference by Enumeration**

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

- We want: $P(Q|e_1 \ldots e_k) = \frac{P(Q,e_1 \ldots e_k)}{\sum_{e_1 \ldots e_k} P(Q,e_1 \ldots e_k)}$

- 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

* Works fine with multiple query variables, too

---

**The Product Rule**

- Sometimes have conditional distributions but want the joint

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

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

- **Example:**

<table>
<thead>
<tr>
<th>$P(W)$</th>
<th>$P(D,W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>sun</td>
<td>0.8</td>
</tr>
<tr>
<td>rain</td>
<td>0.2</td>
</tr>
<tr>
<td>D</td>
<td>W</td>
</tr>
<tr>
<td>wet</td>
<td>sun</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
</tr>
</tbody>
</table>
The Chain Rule

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

\[ P(x_1, x_2, x_3) = \frac{P(x_1, x_2)}{P(x_2|x_1)} \cdot P(x_3|x_1, x_2) \]

- 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:

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

- 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
    - \( P(s|m) = 0.1 \)
    - \( P(s|m) = 0.8 \)
    - \( P(m) = 0.0001 \)

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

  - 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: \( P(G) \)
    - 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:
  \[ P(g|r) \propto P(r|g)P(g) \]
**Ghostbusters, Revisited**

- **P(G):** Prior distribution over ghost location
- **Sensor reading model:** $P(R | G)$

<table>
<thead>
<tr>
<th>$R$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>.8</td>
</tr>
<tr>
<td>$o$</td>
<td>.18</td>
</tr>
<tr>
<td>$y$</td>
<td>.02</td>
</tr>
<tr>
<td>$g$</td>
<td>.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>.2</td>
</tr>
<tr>
<td>$o$</td>
<td>.5</td>
</tr>
<tr>
<td>$y$</td>
<td>.27</td>
</tr>
<tr>
<td>$g$</td>
<td>.3</td>
</tr>
</tbody>
</table>

Bayes' rule:

$$P(G | R) \propto P(R | G) P(G)$$

**Independence**

- Two variables are **independent if**

$$\forall x, y : P(x, y) = P(x) P(y) \quad (1)$$

This says that their joint distribution factors into a product two simpler distributions
- Another form:

$$\forall x, y : P(x | y) = P(x) \quad (2)$$

We write: $X \perp Y$

- Independence is a simplifying modeling assumption
  - **Empirical** joint distributions: at best “close” to independent
  - What could we assume for (Weather, Traffic, Cavity, Toothache)?
Example: Independence?

\[ P(T) \]

\[ \begin{array}{|c|c|}
\hline
T & P \\
\hline
\text{warm} & 0.5 \\
\text{cold} & 0.5 \\
\hline
\end{array} \]

\[ P_1(T, W) \]

\[ \begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{warm} & \text{sun} & 0.4 \\
\text{warm} & \text{rain} & 0.1 \\
\text{cold} & \text{sun} & 0.2 \\
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array} \]

\[ P(W) \]

\[ \begin{array}{|c|c|}
\hline
W & P \\
\hline
\text{sun} & 0.6 \\
\text{rain} & 0.4 \\
\hline
\end{array} \]

Example: Independence

- N fair, independent coin flips:

\[ P(X_1) \]

\[ \begin{array}{|c|c|}
\hline
H & 0.5 \\
T & 0.5 \\
\hline
\end{array} \]

\[ P(X_2) \]

\[ \begin{array}{|c|c|}
\hline
H & 0.5 \\
T & 0.5 \\
\hline
\end{array} \]

\[ \ldots \]

\[ P(X_n) \]

\[ \begin{array}{|c|c|}
\hline
H & 0.5 \\
T & 0.5 \\
\hline
\end{array} \]

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

\[ 2^n \]
Probabilistic Models

- Models describe how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - “All models are wrong; but some are useful.” – George E. P. Box
- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
    - Example: explanation (diagnostic reasoning)
    - Example: prediction (causal reasoning)
    - Example: value of information

For n variables with domain sizes d, joint distribution table with $d^n - 1$ free parameters [recall probabilities sum to one]

Size of representation if we use the chain rule

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

Concretely, counting the number of free parameters accounting for that we know probabilities sum to one:

\[
\frac{(d-1) + d(d-1) + d^2(d-1) + \ldots + d^{n-1}(d-1)}{(d-1)} = \frac{d^n - 1}{d - 1}
\]

[why do both representations have the same number of free parameters?]
Conditional Independence

- \( P(\text{Toothache, Cavity, Catch}) \)
  - 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:
    - \( P(\text{Catch} | \text{Toothache, Cavity}) = P(\text{Catch} | \text{Cavity}) \)
  - Equivalent statements:
    - \( P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity}) \)
    - \( P(\text{Toothache} | \text{Catch, Cavity}) = P(\text{Toothache} | \text{Cavity}) \)
      \( \text{(1)} \)
    - \( P(\text{Toothache}, \text{Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity}) \)
      \( \text{(2)} \)
    - 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:

\[ \forall x, y, z : P(x, y | z) = P(x | z) P(y | z) \]
\[ \forall x, y, z : P(x | z, y) = P(x | z) \]

What about this domain:
- Traffic
- Umbrella
- Raining

What about fire, smoke, alarm?
- Assume A
- Given B
- Therefore C

\( X \perp Y | Z \)