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

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query variables: $Q = Q_1 \ldots Q_r$
  - Hidden variables: $H = H_1 \ldots H_s$
  - We want: $P(Q \mid e_1 \ldots e_k)$
- First, select the entries consistent with the evidence
- Second, sum out $H$ to get joint of Query and evidence:
  $$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_s} P(Q, h_1 \ldots h_s, e_1 \ldots e_k)$$
- 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
  $$P(x \mid y) = \frac{P(x, y)}{P(y)}$$
- Example:

$$P(W) \quad P(D \mid W) \quad P(D, W)$$

<table>
<thead>
<tr>
<th>D</th>
<th>W</th>
<th>P</th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>sun</td>
<td>0.1</td>
<td>wet</td>
<td>sun</td>
<td>0.09</td>
</tr>
<tr>
<td>dry</td>
<td>sun</td>
<td>0.9</td>
<td>dry</td>
<td>sun</td>
<td>0.72</td>
</tr>
<tr>
<td>wet</td>
<td>rain</td>
<td>0.7</td>
<td>wet</td>
<td>rain</td>
<td>0.14</td>
</tr>
<tr>
<td>dry</td>
<td>rain</td>
<td>0.3</td>
<td>dry</td>
<td>rain</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Assignments
- P3 due tonight
- W4 going out tonight

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

- More generally, can always write any joint distribution as an incremental product of conditional distributions
  \[ P(x_1, x_2, x_3) = P(x_1) P(x_2 | x_1) P(x_3 | x_1, x_2) \]
  \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1}) \]
  - 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:
  \[ P(x|y) = \frac{P(y|x)P(x)}{P(y)} \]
  - Why is this always true?
    - 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:
  \[ P(\text{Caused}|\text{Effect}) = \frac{P(\text{Effect}|\text{Caused})P(\text{Caused})}{P(\text{Effect})} \]
- Example:
  - m is meningitis, s is stiff neck
  - \[ P(s|m) = 0.8 \]
  - \[ P(m) = 0.0001 \]
  - \[ P(s) = 0.1 \]

  \[ P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.0 \times 0.0001}{0.1} = 0.00008 \]
  - 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:
    \[ P(G|r) \propto P(r|G)P(G) \]

Independence

- Two variables are independent if:
  \[ \forall x, y : P(x, y) = P(x)P(y) \]
  - This says that their joint distribution factors into a product two simpler distributions
  - Another form:
    \[ \forall x, y : P(x|y) = P(x) \]
  - We write: \( X \perp \!
  \perp Y \)
  - 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?

\[
P(T) = \begin{array}{c|cc}
  & P & P \\
  \text{warm} & 0.5 & 0.4 \\
  \text{cold} & 0.5 & 0.6 \\
\end{array}
\]

\[
P(W) = \begin{array}{c|cc}
  & P & P \\
  \text{sun} & 0.4 & 0.6 \\
  \text{rain} & 0.6 & 0.4 \\
\end{array}
\]

Example: Independence

- N fair, independent coin flips:

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

\[
P(X_1, X_2, \ldots, X_n) = \prod P(x_i | x_1 \ldots x_{i-1})
\]

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

Conditional Independence

- P(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:
    \[P(\text{Catch} | \text{Toothache}, \text{Cavity}) = P(\text{Catch} | \text{Cavity})\]
  - 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

Conditional Independence

- 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?