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 variable: $Q$.
  - Hidden variables: $H_1, \ldots, H_r$.
- 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 condition.

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

**Example:**

$P(W)$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>rain</td>
<td>2</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

$P(D|W)$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>dry</td>
<td>1</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

$P(D,W)$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>wet</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>dry</td>
<td>1</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>
The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions
  \[ P(x_1, x_2, x_3) = \frac{P(x_1)P(x_2|x_1)P(x_3|x_2, x_1)}{P(x_2|x_1)} \]

- Often one conditional is tricky but the other one is simple
  - Let’s build one conditional from its reverse

- Why is this always true?

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
  - Example given:
    \[ P(s|m) = 0.1 \]
    \[ P(m) = 0.8 \]
    \[ P(s) = 0.0001 \]

- Note: posterior probability is not necessarily the same as prior
- Note: you should get stiff necks checked out! Why?

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 so 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!

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 | 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)

| P(R | G) | P(R | 0) | P(R | 1) | P(R | 2) | P(R | 3) |
|-------|--------|--------|--------|--------|
| 0.15  | 0.25   | 0.27   | 0.22   | 0.12   |

- 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: \( X \perp Y \)

- 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?

\[
P(T) = \begin{pmatrix}
T & W \\
\text{warm} & 0.5 \\
\text{cold} & 0.5
\end{pmatrix}
\]

\[
P(W) = \begin{pmatrix}
W & P \\
\text{sun} & 0.6 \\
\text{rain} & 0.4
\end{pmatrix}
\]

Example: Independence

- N fair, independent coin flips:

\[
P(X_1) = \begin{pmatrix}
H & 0.5 \\
T & 0.5
\end{pmatrix}
\]

\[
P(X_2) = \begin{pmatrix}
H & 0.5 \\
T & 0.5
\end{pmatrix}
\]

\[
P(X_n)
\]

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

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:

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

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

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

Conditional independence is very rare (why?)

- What about this domain:
  - Traffic
  - Umbrella
  - Raining
- What about fire, smoke, alarm?