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
Model for Ghostbusters

- Reminder: ghost is hidden, sensors are noisy
- $T$: Top sensor is red
  - $B$: Bottom sensor is red
  - $G$: Ghost is in the top
- Queries:
  - $P(+g) = ??$
  - $P(+g | +t) = ??$
  - $P(+g | +t, -b) = ??$
- Problem: joint distribution too large / complex

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>B</th>
<th>G</th>
<th>$P(T,B,G)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>+b</td>
<td>+g</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>+t</td>
<td>+b</td>
<td>-g</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>+t</td>
<td>-b</td>
<td>+g</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>+t</td>
<td>-b</td>
<td>-g</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>+b</td>
<td>+g</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>+b</td>
<td>-g</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>-b</td>
<td>+g</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>-b</td>
<td>-g</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

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

- N fair, independent coin flips:

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

\[
\begin{array}{c|c}
  h & 0.5 \\
  t & 0.5
\end{array}
\quad
\begin{array}{c|c}
  h & 0.5 \\
  t & 0.5
\end{array}
\quad \ldots \\
\begin{array}{c|c}
  h & 0.5 \\
  t & 0.5
\end{array}
\]

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

\[
2^n
\]

---

Example: Independence?

\[
P(T)
\]

\[
P_1(T, W)
\]

\[
<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>warm</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>

\[
P_2(T, W)
\]

\[
P(W)
\]

\[
<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>0.4</td>
</tr>
<tr>
<td>warm</td>
<td>sun</td>
</tr>
<tr>
<td>warm</td>
<td>rain</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
</tr>
</tbody>
</table>
Conditional Independence

- \( P(\text{Toothache}, \text{Cavity}, \text{Catch}) \)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  \[ P(+\text{catch} \mid +\text{toothache}, +\text{cavity}) = P(+\text{catch} \mid +\text{cavity}) \]

- The same independence holds if I don’t have a cavity:
  \[ P(+\text{catch} \mid +\text{toothache}, -\text{cavity}) = P(+\text{catch} \mid -\text{cavity}) \]

- Catch is conditionally independent of Toothache given Cavity:
  \[ P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \]

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

\[
\forall x, y, z : P(x, y \mid z) = P(x \mid z)P(y \mid z) \\
X \perp Y \mid Z
\]

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

- What about fire, smoke, alarm?
The Chain Rule

\[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots \]

- Trivial decomposition:
  \[ P(\text{Traffic, Rain, Umbrella}) = \]
  \[ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain, Traffic}) \]

- With assumption of conditional independence:
  \[ P(\text{Traffic, Rain, Umbrella}) = \]
  \[ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}) \]

- Bayes’ nets / graphical models help us express conditional independence assumptions

Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is
- That means, the two sensors are conditionally independent, given the ghost position
- \( T: \) Top square is red
  \( B: \) Bottom square is red
  \( G: \) Ghost is in the top

- Givens:
  \[ P(\text{+} g) = 0.5 \]
  \[ P(\text{+} t | \text{+} g) = 0.8 \]
  \[ P(\text{+} t | \text{-} g) = 0.4 \]
  \[ P(\text{+} b | \text{+} g) = 0.4 \]
  \[ P(\text{+} b | \text{-} g) = 0.8 \]

\[
P(T, B, G) = P(G)P(T|G)P(B|G)
\]

<table>
<thead>
<tr>
<th>T</th>
<th>B</th>
<th>G</th>
<th>( P(T, B, G) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>+b</td>
<td>+g</td>
<td>0.16</td>
</tr>
<tr>
<td>+t</td>
<td>+b</td>
<td>−g</td>
<td>0.16</td>
</tr>
<tr>
<td>+t</td>
<td>−b</td>
<td>+g</td>
<td>0.24</td>
</tr>
<tr>
<td>+t</td>
<td>−b</td>
<td>−g</td>
<td>0.04</td>
</tr>
<tr>
<td>−t</td>
<td>+b</td>
<td>+g</td>
<td>0.04</td>
</tr>
<tr>
<td>−t</td>
<td>+b</td>
<td>−g</td>
<td>0.24</td>
</tr>
<tr>
<td>−t</td>
<td>−b</td>
<td>+g</td>
<td>0.06</td>
</tr>
<tr>
<td>−t</td>
<td>−b</td>
<td>−g</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Bayes’ Nets: Big Picture

- Two problems with using full joint distribution tables as our probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes’ nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we’ll be vague about how these interactions are specified

Example Bayes’ Net: Insurance
Example Bayes’ Net: Car

Graphical Model Notation

- **Nodes**: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)

- **Arcs**: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)

- For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Coin Flips

- N independent coin flips

\[ X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n \]

- No interactions between variables: absolute independence

Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

- Model 2: rain causes traffic

- Why is an agent using model 2 better?
Example: Traffic II

- Let’s build a causal graphical model

- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayes’ Net Semantics

- Let’s formalize the semantics of a Bayes’ net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents’ values
    \[ P(X|a_1 \ldots a_n) \]
  - CPT: conditional probability table
  - Description of a noisy “causal” process

* A Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BNs

- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
    \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i)) \]
  - Example:
    \[ P(\text{+cavity}, \text{+catch}, \neg\text{toothache}) \]
- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies
Example: Coin Flips

\( P(h, h, t, h) = \)

Only distributions whose variables are absolutely independent can be represented by a Bayes’ net with no arcs.

Example: Traffic

\( P(+r, -t) = \)

\begin{align*}
P(R) & \begin{array}{c|c} +r & 1/4 \\ \hline
-\overline{r} & 3/4 \\ \hline
\end{array} \\
P(T|R) & \begin{array}{c|c|c} +r & +t & 3/4 \\ \hline
\overline{r} & +t & 1/2 \\
\hline
-\overline{r} & -t & 1/4 \\
\hline
\end{array} \\
\end{align*}
**Example: Alarm Network**

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>−b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

**Example: Traffic**

- **Causal direction**

| P(R) | P(T|R) |
|------|--------|
| r    | t 3/4  |
| r    | t 1/4  |
| −r   | t 3/4  |
| −r   | t 1/4  |
| t    | −t 1/4 |
| t    | −t 1/2 |
| −t   | −t 1/2 |
| −t   | −t 6/16 |
| −t   | −t 6/16 |

<table>
<thead>
<tr>
<th>P(R)</th>
<th>P(T, R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>t 3/16</td>
</tr>
<tr>
<td>r</td>
<td>t 1/16</td>
</tr>
<tr>
<td>−r</td>
<td>t 3/16</td>
</tr>
<tr>
<td>−r</td>
<td>t 1/16</td>
</tr>
<tr>
<td>t</td>
<td>−t 1/4</td>
</tr>
<tr>
<td>t</td>
<td>−t 1/2</td>
</tr>
<tr>
<td>−t</td>
<td>−t 1/2</td>
</tr>
<tr>
<td>−t</td>
<td>−t 6/16</td>
</tr>
<tr>
<td>−t</td>
<td>−t 6/16</td>
</tr>
</tbody>
</table>
Example: Reverse Traffic

- Reverse causality?

\[
P(T) \hspace{1cm} P(T, R)
\begin{array}{c|c|c|}
  \text{t} & 9/16 & r \\ 
  \neg t & 7/16 & \neg r \\
\end{array}
\begin{array}{c|c|c|}
  r & t & 3/16 \\ 
  r & \neg t & 1/16 \\
\end{array}
\begin{array}{c|c|c|}
  \neg t & r & 1/3 \\ 
  \neg t & \neg r & 2/3 \\
\end{array}
\begin{array}{c|c|c|}
  \neg r & t & 6/16 \\ 
  \neg r & \neg t & 6/16 \\
\end{array}
\]

Causality?

- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. consider the variables Traffic and Drips
  - End up with arrows that reflect correlation, not causation

- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence
Bayes’ Nets

- So far: how a Bayes’ net encodes a joint distribution
- Next: how to answer queries about that distribution
  - Key idea: conditional independence
  - Today: assembled BNs using an intuitive notion of conditional independence as causality
  - Next: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)