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

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?

<table>
<thead>
<tr>
<th>P(T)</th>
<th>P(W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>P</td>
</tr>
<tr>
<td>hot</td>
<td>0.5</td>
</tr>
<tr>
<td>cold</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Example: Independence

- N fair, independent coin flips:
  - $P(X_1)$
  - $P(X_2)$
  - ... $P(X_n)$

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:
  - $P(+\text{catch} | +\text{toothache}, +\text{cavity}) = P(+\text{catch} | +\text{cavity})$
- 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.
- $X$ is conditionally independent of $Y$ given $Z$ 
  - $X \perp Y | Z$
  - if and only if:
    - $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$
  - or, equivalently, if and only if
    - $\forall x, y, z : P(x|z, y) = P(x|z)$
Conditional Independence

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

Conditional Independence and the Chain Rule

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

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

- T, B, G
  - 0.50

- \[ \begin{array}{ccc}
    +t & +b & +g \\
    0.16 \\
    +t & +b & -g \\
    0.16 \\
    +t & -b & +g \\
    0.24 \\
    +t & -b & -g \\
    0.04 \\
    -t & +b & +g \\
    0.04 \\
    -t & +b & -g \\
    0.24 \\
    -t & -b & +g \\
    0.06 \\
    -t & -b & -g \\
    0.06
  \end{array} \]
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**

**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

- No interactions between variables: absolute independence
Example: Alarm Network

- **Variables**
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

Bayes’ Net Semantics

- 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:
  \[ P(+\text{cavity}, +\text{catch}, -\text{toothache}) \]
Probabilities in BNs

- Why are we guaranteed that setting
  \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i)) \]
  results in a proper joint distribution?

- Chain rule (valid for all distributions):
  \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1}) \]

- Assume conditional independences:
  \[ P(x_i | x_1 \ldots x_{i-1}) = P(x_i | \text{parents}(X_i)) \]
  → Consequence: \[ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i)) \]

- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies

Example: Coin Flips

\[
\begin{array}{c|c|c|c}
\text{X}_1 & \text{X}_2 & \ldots & \text{X}_n \\
\hline
\text{h} & 0.5 & \text{h} & 0.5 & \ldots & \text{h} & 0.5 \\
\text{t} & 0.5 & \text{t} & 0.5 & \ldots & \text{t} & 0.5 \\
\end{array}
\]

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

Example: Traffic

\[
P(R) \\
\begin{array}{c|c}
\text{r} & 0.25 \\
\text{t} & 0.75 \\
\end{array}
\]

\[
P(T | R) \\
\begin{array}{c|c|c}
\text{r} & \text{t} & \text{p} \\
\hline
\text{+} & 0.67 & 0.17 \\
\text{-} & 0.33 & 0.83 \\
\end{array}
\]

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

Example: Alarm Network

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.
Example: Traffic

- Causal direction

<table>
<thead>
<tr>
<th></th>
<th>P(R)</th>
<th></th>
<th>P(T, R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>1/4</td>
<td>-r</td>
<td>3/4</td>
</tr>
<tr>
<td>-r</td>
<td>3/4</td>
<td>+r</td>
<td>1/4</td>
</tr>
</tbody>
</table>

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
  - Today:
    - First assembled BNs using an intuitive notion of conditional independence as causality
    - Then saw that key property is conditional independence
  - Main goal: answer queries about conditional independence and influence

- After that: how to answer numerical queries (inference)