Announcements

- Written assignment 2
  - Due this Thursday in lecture!
  - Solutions posted on Friday: late assignments will not be accepted after solutions are posted

- Midterm is one week from Thursday (3/19)
  - One page (2 sides) of notes & calculator allowed
  - Review session on...

- Course contest announced on Thursday

---

**Bayes’ Nets**

- A Bayes’ net is an efficient encoding of a probabilistic model of a domain

- Questions we can ask:
  - Inference: given a fixed BN, what is $P(X | e)$?
  - Representation: given a fixed BN, what kinds of distributions can it encode?
  - Modeling: what BN is most appropriate for a given domain?

---

**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

---

**Building the (Entire) Joint**

- We can take a Bayes’ net and build any entry from the full joint distribution it encodes
  $$ P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \text{parents}(X_i)) $$

  - Typically, there’s no reason to build ALL of it
  - We build what we need on the fly

  - To emphasize: every BN over a domain implicitly defines a joint distribution over that domain, specified by local probabilities and graph structure

---

**Example: Alarm Network**

$$ \prod_{i} P(X_i | \text{Parents}(X_i)) = P(B) \cdot P(E) \cdot P(A | B, E) \cdot P(J | A) \cdot P(M | A) $$
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>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alarm</th>
<th>B</th>
<th>E</th>
<th>A</th>
<th>P(A,B,E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>John calls</td>
<td>b</td>
<td>e</td>
<td>a</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-e</td>
<td>-a</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-e</td>
<td>e</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-e</td>
<td>a</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-e</td>
<td>-a</td>
<td>0.71</td>
</tr>
<tr>
<td>Mary calls</td>
<td>b</td>
<td>e</td>
<td>a</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>-e</td>
<td>a</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Bayes’ Nets So Far

- We now know:
  - What is a Bayes’ net?
  - What joint distribution does a Bayes’ net encode?
- Next: properties of that joint distribution (independence)
  - Key idea: conditional independence
  - Last class: assembled BNs using an intuitive notion of conditional independence as causality
  - Today: formalize these ideas
  - Main goal: answer queries about conditional independence and influence
- Tomorrow: how to compute posteriors quickly (inference)

Size of a Bayes’ Net

- How big is a joint distribution over N Boolean variables? $2^N$
- How big is an N-node net if nodes have up to k parents? $O(N \cdot 2^{k+1})$
- Both give you the power to calculate $P(X_1, X_2, \ldots, X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Conditional Independence

- Reminder: independence
  - X and Y are independent if $\forall x, y \ P(x, y) = P(x)P(y)$ $\rightarrow$ $X \perp Y$
- X and Y are conditionally independent given Z
  - $\forall x, y, z \ P(x, y|z) = P(x|z)P(y|z)$ $\rightarrow$ $X \perp Y | Z$
- (Conditional) independence is a property of a distribution

Example: Independence

- For this graph, you can fiddle with it (the CPTs) all you want, but you won’t be able to represent any distribution in which the flips are dependent!

<table>
<thead>
<tr>
<th>X_1</th>
<th>X_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>t</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X_1, \perp X_2</th>
</tr>
</thead>
</table>

Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution
Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can calculate using algebra (really tedious)
  - If no, can prove with a counter example

- Example:

```
X ---- Y ---- Z
```

- Question: are X and Z independent?
  - Answer: not necessarily, we've seen examples otherwise:
    - low pressure causes rain, which causes traffic;
    - X can influence Z, Z can influence X (via Y)
  - Addendum: they could be independent: how?

Common Cause

- Another basic configuration: two effects of the same cause
  - Are X and Z independent?
  - Are X and Z independent given Y?

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

- Observing the cause blocks influence between effects.

Common Effect

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
  - Are X and Z independent given Y?

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

- Observing the effect enables influence between effects.

The General Case

- Any complex example can be analyzed using these three canonical cases
  - General question: in a given BN, are two variables independent (given evidence)?
  - Solution: analyze the graph

Causal Chains

- This configuration is a "causal chain"

```
X ---- Y ---- Z
```

- Is X independent of Z given Y?

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

- Evidence along the chain "blocks" the influence

The General Case

- Any complex example can be analyzed using these three canonical cases
  - General question: in a given BN, are two variables independent (given evidence)?
  - Solution: analyze the graph

Reachability

- Recipe: shade evidence nodes
  - Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
  - Almost works, but not quite
    - Where does it break?
      - Answer: the v-structure at T doesn’t count as a link in a path unless “active”
Reachability (D-Separation)

- **Question:** Are X and Y conditionally independent given evidence variables \{Z\}?
  - Look for "active paths" from X to Y
  - No active paths = independence!

- **A path is active if each triple is either a:**
  - Causal chain A → B → C where B is unobserved (either direction)
  - Common cause A ← B → C where B is unobserved
  - Common effect (aka v-structure) A → B ← C where B or one of its descendents is observed

Active Triples

Inactive Triples

Example

- **Variables:**
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad

- **Questions:**
  - \( T \perp D \)
  - \( T \perp D | R \)
  - \( T \perp D | R, S \)

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
  - 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 only guaranteed to encode conditional independence

Example: Traffic

- **Basic traffic net**
- Let’s multiply out the joint

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

| \( P(T | R) \) | \( P(T | \neg R) \) |
|---|---|
| \( r \) | \( t \) | 3/4 | \( r \) | \( t \) | 1/2 |
| \( \neg r \) | 1/2 | \( \neg r \) | 1/2 |
Example: Reverse Traffic

- Reverse causality?

<table>
<thead>
<tr>
<th></th>
<th>P(T)</th>
<th></th>
<th>P(T, R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>t</td>
<td>3/16</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>t</td>
<td>1/16</td>
</tr>
<tr>
<td>R</td>
<td>r</td>
<td>t</td>
<td>5/16</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>t</td>
<td>5/16</td>
</tr>
<tr>
<td>T</td>
<td>t</td>
<td>r</td>
<td>3/16</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>r</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>r</td>
<td>5/16</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>r</td>
<td>5/16</td>
</tr>
<tr>
<td></td>
<td>-r</td>
<td>t</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>-r</td>
<td>t</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>-r</td>
<td>-r</td>
<td>5/16</td>
</tr>
<tr>
<td></td>
<td>-r</td>
<td>-r</td>
<td>5/16</td>
</tr>
<tr>
<td></td>
<td>-t</td>
<td>r</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>-t</td>
<td>r</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>-t</td>
<td>-r</td>
<td>1/16</td>
</tr>
<tr>
<td></td>
<td>-t</td>
<td>-r</td>
<td>1/16</td>
</tr>
</tbody>
</table>

Example: Coins

- Extra arcs don’t prevent representing independence, just allow non-independence

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th></th>
<th>X2</th>
<th></th>
<th>X1</th>
<th></th>
<th>X2</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>h</td>
<td>0.5</td>
<td>h</td>
<td>0.5</td>
<td>h</td>
<td>h</td>
<td>0.5</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>0.5</td>
<td>t</td>
<td>0.5</td>
<td>h</td>
<td>t</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Adding arcs isn’t wrong, it’s just inefficient

Example: Alternate Alarm

If we reverse the edges, we make different conditional independence assumptions

To capture the same joint distribution, we have to add more edges to the graph

Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence assumptions
- A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution