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 \( P(X \mid e) \)?
  - 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
      \[
      P(X \mid a_1 \ldots a_n)
      \]
  - CPT: conditional probability table
  - Description of a noisy “causal” process

\[ A \text{ Bayes net} = \text{Topology (graph)} + \text{Local Conditional Probabilities} \]
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>

John calls

| A  | J  | P(J|A) |
|----|----|------|
| a  | j  | 0.9  |
| a  | ¬j | 0.1  |
| ¬a | j  | 0.05 |
| ¬a | ¬j | 0.95 |

Mary calls

| A  | M  | P(M|A) |
|----|----|------|
| a  | m  | 0.7  |
| a  | ¬m | 0.3  |
| ¬a | m  | 0.01 |
| ¬a | ¬m | 0.99 |

Alarm

| B | E | A | P(A|B,E) |
|---|---|---|---------|
| b | e | a | 0.95    |
| b | e | ¬a| 0.05   |
| b | ¬e| a | 0.94   |
| b | ¬e| ¬a| 0.06  |
| ¬b| e | a | 0.29   |
| ¬b| e | ¬a| 0.71  |
| ¬b| ¬e| a | 0.001 |
| ¬b| ¬e| ¬a| 0.999 |

Earthquake

| A  | J  | P(J|A) |
|----|----|------|
| a  | j  | 0.9  |
| a  | ¬j | 0.1  |
| ¬a | j  | 0.05 |
| ¬a | ¬j | 0.95 |

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? 
  \(O(N \times 2^{k+1})\)

- 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)
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)

Conditional Independence

- Reminder: independence
  - $X$ and $Y$ are independent if
    \[ \forall x, y \quad P(x, y) = P(x)P(y) \rightarrow X \perp Y \]
  - $X$ and $Y$ are conditionally independent given $Z$
    \[ \forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \rightarrow X \perp Y|Z \]

- (Conditional) independence is a property of a distribution
Example: Independence

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

\[ X_1 \quad X_2 \]

\[
\begin{array}{c|c}
P(X_1) & P(X_2) \\
\hline
h & h \\
t & t \\
0.5 & 0.5 \\
\end{array}
\]

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:

\[ X \rightarrow Y \rightarrow Z \]

- 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?

Causal Chains

- This configuration is a “causal chain”

\[ X \rightarrow Y \rightarrow Z \]

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

- 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
**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?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - No: seeing traffic puts the rain and the ballgame in competition as explanation?
  - This is backwards from the other cases
    - 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

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 \( A \rightarrow B \rightarrow C \) where B is unobserved (either direction)
  - Common cause \( A \leftarrow B \rightarrow C \) where B is unobserved
  - Common effect (aka v-structure) \( A \rightarrow B \leftarrow C \) where B or one of its descendents is observed

Example

\[
A \perp W \quad \text{Yes}
\]

\[
A \perp W | R
\]
### Example

\[ L \perp T' | T \quad \text{Yes} \]
\[ L \perp B \quad \text{Yes} \]
\[ L \perp B | T \]
\[ L \perp B | T' \]
\[ L \perp B | T, R \quad \text{Yes} \]

### Example

- **Variables:**
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad
- **Questions:**
  \[ T \perp D \]
  \[ T \perp D | R \quad \text{Yes} \]
  \[ T \perp D | R, S \]
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

\[
\begin{array}{c|c}
R & P(R) \\
\hline
r & 1/4 \\
\neg r & 3/4 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
R & T & P(T|R) \\
\hline
r & t & 3/4 \\
\neg r & t & 1/4 \\
\neg r & \neg t & 1/2 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
T & P(T,R) \\
\hline
r & t & 3/16 \\
r & \neg t & 1/16 \\
\neg r & t & 6/16 \\
\neg r & \neg t & 6/16 \\
\end{array}
\]
Example: Reverse Traffic

- Reverse causality?

```
P(T)
+---+----+
|   | t  | 9/16 |
|---+----+-----|
| ¬t| 7/16|
+---+----+

P(R|T)
+---+---+-----+
|   | r  | 1/3  |
|---+----+-----|
| ¬r| 2/3 |
+---+---+-----+
| ¬t| r  | 1/7  |
| ¬r| 6/7 |
+---+---+-----+

P(T, R)
+---+---+-----+
|   | r  | 3/16 |
+---+----+-----|
| r  | ¬t | 1/16 |
+---+----+-----|
| ¬r| t | 6/16 |
+---+----+-----|
| ¬r| ¬t | 6/16 |
+---+----+-----+
```

Example: Coins

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

```
P(X1)
+---+---+
| h | 0.5|
+---+---+
| t | 0.5|
+---+---+

P(X2)
+---+---+
| h | 0.5|
+---+---+
| t | 0.5|
+---+---+

P(X1)
+---+---+
| h | 0.5|
+---+---+
| t | 0.5|
+---+---+

P(X2|X1)
+---+---+---+
| h | h | 0.5 |
| t | h | 0.5 |
+---+---+---+
| h | t | 0.5 |
+---+---+---+
| t | t | 0.5 |
+---+---+---+
```

- Adding arcs isn’t wrong, it’s just inefficient
Changing Bayes’ Net Structure

- The same joint distribution can be encoded in many different Bayes’ nets
  - Causal structure tends to be the simplest

- Analysis question: given some edges, what other edges do you need to add?
  - One answer: fully connect the graph
  - Better answer: don’t make any false conditional independence assumptions

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
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