Lecture 20: Bayes Nets
Prof. Louis Steinberg

CS 530: Artificial Intelligence
• The state of a single situation / domain may be represented by a set of random variables.

Joint Probability

• Joint probability distribution (or just “the joint”): probability distribution over all possible sets of values for the state variables.
### Example of Joint Probability Distribution

#### Minesweeper

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>MinesLeft = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0</td>
</tr>
</tbody>
</table>
Joint Probability Distribution

- Choose a set of random variables: $X_1, \ldots, X_n$.
- An atomic event $(X_1=x_1, \ldots, X_n=x_n)$ is an assignment of exactly one value to each variable.
- The joint probability distribution function assigns a probability $P(X_1=x_1, \ldots, X_n=x_n)$ to each atomic event.
- Encode as a table:
  - Size $2^N$ for boolean random variables.
  - Size $D^N$ for general random variables.

Joint Probability Distribution
Difficulties:
- Most joint entries are independent of each other, or at least conditionally independent.
- Constructing the joint requires acquiring $D^N$ numbers from an expert.
- Computation with the joint requires lots of memory and CPU time.

Opportunity:
- The world is nearly decomposable. (Most things don't interact with each other.)
- The joint is nearly independent. (The world is nearly decomposable.)
- Computation with the joint requires acquiring $D^N$ numbers from an expert.
- Constructing the joint requires acquiring $D^N$ numbers from an expert.

Using the Joint Probability Distribution
Belief Networks

- Structure encodes conditional independence.
  - Each arc represents a direct influence of one variable on another.
- Each node represents a random variable.
- A directed acyclic graph (DAG).
- A graphical representation of the joint probability distribution.
- A graphical representation of the joint probability distribution attempts to overcome difficulties of using the joint probability distribution.
A Typical Belief Network
Semantics of Belief Networks

- A belief network encodes conditional probabilities.
- The table at the node for variable $X_i$ stores the conditional probability:
  - for each combination of values of $\forall j \in \text{Parents}(i)$. 
- For each combination of values of $\forall j \in \text{Parents}(i)$ and values of $X_i$.
Conditional Probabilities
Quantities in the Network

- $P(B) = 0.001$
- $P(E) = 0.002$
- $P(A | B \land E) = 0.95$
- $P(A | B \land \sim E) = 0.94$
- $P(A | \sim B \land E) = 0.29$
- $P(A | \sim B \land \sim E) = 0.001$
- $P(J | A) = 0.90$
- $P(J | \sim A) = 0.05$
- $P(M | A) = 0.70$
- $P(M | \sim A) = 0.01$
Boolean Case

Let \( A \) be an atomic event.

Belief Net

Extracting the Joint from a Belief Net

\[
p(A) = \prod_{i=1}^{n} p(X_i | \text{Parents}(i))
\]

It follows that:

\[
A = \bigwedge_{1}^{n} \neg X_i \lor \bigwedge_{1}^{n} X_i
\]
Conditional Independence Property:
Each variable is conditionally independent of its predecessors given its parents.
Constructing a Belief Network

1. Choose a set of variables.
2. Choose an ordering of the variables.
3. While (Some variables are left)
   a. Choose a variable, $X_i$.
   b. Add a node for $X_i$ to the network.
   c. Set $\text{Parents}(i)$ to be the minimal set of nodes already in the network satisfying the conditional independence property.
   d. Define the conditional probability table for node $X_i$.
4. Choose a variable, $X_{i+1}$.
5. While (Some variables are left)
Network Structure Depends on Order of Introduction
Illustration of Conditional Independence

- RADIO and GAS independent given nothing.
- RADIO and GAS independent given BATTERY.
- RADIO and GAS independent given IGNITION.
- RADIO and GAS independent given nothing.
Patterns of Reasoning Handled by Belief Networks

- **Diagnostic Inference:**
  - Given JohnCalls infer \( P(\text{Burglary}|\text{JohnCalls}) \)

- **Causal Inference:**
  - Given Burglary infer \( P(\text{JohnCalls}|\text{Burglary}) \)

- **Intercausal Inference:**
  - Given Alarm and Earthquake infer \( P(\text{Burglary}|\text{Alarm} \land \text{Earthquake}) \)

- **Mixed Inference:**
  - Given Burglary infer \( P(\text{Alarm}|\text{Burglary}) \)

- **Causal Inference:**
  - Given JohnCalls infer \( P(\text{Burglary}|\text{JohnCalls}) \)

- **Diagnostic Inference:**
  - Given Burglary infer \( P(\text{JohnCalls}|\text{Burglary}) \)
Terminology

- Let "B" be "Burglary".
- Let "~E" be "Earthquake".
- Let "~A" be "Alarm".
- Let "~M" be "Mary Calls".
- Let "~J" be "John Calls".
Probability that there was a burglary given that the alarm goes off:

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

Probability the alarm goes off given that there is a burglary:

$$P(A|B) = P(A \land E|B) + P(A \land \neg E|B) = P(A|E \land B) P(E|B) + P(A|\neg E \land B) P(\neg E|B)$$

Belief Network:

- Probability the alarm goes off given that there is a burglary:
  $$P(A|B) = \frac{P(A \land E|B) + P(A \land \neg E|B)}{P(B)}$$
- The alarm goes off:
  $$P(B|A) = \frac{P(A \land E|B) P(E|B)}{P(A)}$$
Probability that the alarm goes off:

\[
\sum_{B, E} \left( P(A | B, E) P(B) P(E) + P(A | B, \neg E) P(B) P(\neg E) + P(A | \neg B, E) P(\neg B) P(E) + P(A | \neg B, \neg E) P(\neg B) P(\neg E) \right)
\]

Belief Network

Sample Calculations with the
Sample Calculations with the Belief Network

Probability that there is a burglary and MaryCalls:

\[ P(B \land M) = P(B) P(M \land A | B) + P(B) P(M \land \sim A | B) \]

\[ = P(B) P(M | A) P(A | B) + P(B) P(M | \sim A) P(\sim A | B) \]