(More) Efficient CPDs
CPD Tables

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{X_i}) \]

Factors

\[ X_{i1} \quad X_{i2} \quad X_{i3} \quad X_{i4} \]

\[ X_i \]
Size of a Full CPD Table

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{X_i}) \]

Factors

Example: binary variables, \( k \) parents

\[ |CPD(X_i)| = \left[ \prod_{X_j \in Pa_{X_i}} |Val(X_j)| \right] (|Val(X_i)| - 1) \]

STAT 598L: Probabilistic Graphical Models (Efficient CPDs)
How to Reduce the Sizes of CPDs

• Functional dependence of values (deterministic)

• Functional dependence of probabilities
  – Sigmoid (logistic regression)
  – Noisy-or

• Context-specific
  – Tree-CPDs
  – Rule-based
  – Graphs-based CPDs
General View: Function of Inputs

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{X_i}) \]

- Conditional probability distribution inside each factor = function from parents to a binomial distribution

\[ Pa_{X_i} = (Y_1, \ldots, Y_m) \times Val(Y_1) \times \cdots \times Val(Y_m) \]

\{P(X_i=0), P(X_i=1), \ldots, P(X_i=k-1)\}
Functional Dependence

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{X_i}) \]

- **Value** of a random variable each factor = **function** from parents to a binomial distribution
  - Examples: xor, nand
Possible Effect on d-Separation

• Can deterministic CPDs affect d-separation?
  – Yes!
Noisy-Or

• Independent causal mechanisms
  – Each cause is enough to trigger the effect
    • Either earthquake or burglary triggers alarm
    • The effect is not triggered only if none of the causes are active
  – Equivalent to logical or
Noisy-Or

• Suppose causes only trigger the effect with some probability ($p_i$)

\[
P(\neg \text{Alarm} \mid \neg \text{Burglary}, \neg \text{Earthquake}) = 1 - p_B = 0.05
\]
\[
P(\neg \text{Alarm} \mid \neg \text{Burglary}, \text{Earthquake}) = 1 - p_E = 0.1
\]
\[
P(\neg \text{Alarm} \mid \text{Burglary}, \text{Earthquake}) = (1 - p_B)(1 - p_E) = 0.005
\]
\[
P(\neg \text{Alarm} \mid \neg \text{Burglary}, \neg \text{Earthquake}) = 1
\]
Noisy-Or

• More generally (for multivariate binary variables)

\[
P(X_i = 0 \mid X_{i1}, \ldots, X_{ik}) = (1 - p_0) \prod_{i : X_{i1} = 1} (1 - p_i)
\]

\[
P(X_i = 1 \mid X_{i1}, \ldots, X_{ik}) = 1 - (1 - p_0) \prod_{i : X_{i1} = 1} (1 - p_i)
\]
Context-Based CPDs

• What if a particular value of one variable makes some other variables irrelevant?
  – How to encode that?
Tree-CPDs

• Represent as a binary tree

| Y | Z | W | P(X = 1 | Y, Z, W) |
|---|---|---|-----------------|
| 1 | 1 | 1 | 0.85            |
| 1 | 1 | 0 | 0.3             |
| 1 | 0 | 1 | 0.85            |
| 1 | 0 | 0 | 0.3             |
| 0 | 1 | 1 | 0.15            |
| 0 | 1 | 0 | 0.3             |
| 0 | 0 | 1 | 0.5             |
| 0 | 0 | 0 | 0.3             |
Tree-CPDs

• Select order of the nodes to make the tree sparse

\[
\begin{array}{c|c|c|c|c|c}
Y & Z & W & P(X = 1 | Y, Z, W) \\
\hline
1 & 1 & 1 & 0.85 \\
1 & 1 & 0 & 0.3 \\
1 & 0 & 1 & 0.85 \\
1 & 0 & 0 & 0.3 \\
0 & 1 & 1 & 0.15 \\
0 & 1 & 0 & 0.3 \\
0 & 0 & 1 & 0.5 \\
0 & 0 & 0 & 0.3 \\
\end{array}
\]
Multiplexer

• Special type of Tree-CPD
  – Decides which variable to use in a CPD

\[
P(X \mid A, X_1, \ldots, X_k)
\]
Generalization: Decision Graphs

• What if a tree contains many branches?
  – May be able to merge them
Another Generalization: Rules

• What if the tree is not very efficient?
  – Just record the partition as a set of rules

- Probabilistic Graphical Models (Efficient CPDs)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>( P(X = 1 \mid Y, Z, W) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.85</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.30</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.35</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.30</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.70</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Another Generalization: Rules

\[(Y = 0 \land Z = 1) \lor (Y = 1 \land Z = 1 \land W = 0) \Rightarrow P(X = 1) = 0.30\]
\[(Y = 1 \land Z = 1 \land W = 1) \lor (Y = 0 \land Z = 0 \land W = 0) \Rightarrow P(X = 1) = 0.85\]
\[(Y = 1 \land Z = 0 \land W = 1) \Rightarrow P(X = 1) = 0.75\]
\[(Y = 1 \land Z = 0 \land W = 0) \Rightarrow P(X = 1) = 0.35\]
\[(Y = 0 \land Z = 0 \land W = 1) \Rightarrow P(X = 1) = 0.5\]
Possible Effect on d-Separation

• Can context-based CPDs affect d-separation?
  – Yes!