Artificial Intelligence
CS 6364

Professor Dan Moldovan

Section 13
Bayesian Networks
Bayesian Networks

Directed graphs in which each node is annotated with quantitative probability information. The full specification is:

1. A set of random variables makes up the nodes of the network. Variables may be discreet or continuous.
2. A set of directed links or arrows connects pairs of nodes. If there is an arrow from node X to node Y, X is said to be a parent to Y.
3. The graph has no directed cycles (it is a directed acyclic graph, or DAG).
Example 1

- Dental world: Variables: Toothache, Cavity, Catch, Weather

Intuitive idea: an arrow between X and Y means that X has direct influence on Y
Example 2

New burglar alarm is installed at home. It is fairly reliable at detecting a burglary, but it also responds on occasion to minor earthquakes. There are 2 neighbors: John and Mary, who have promised to call you at work when they hear the alarm.

John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm calls and calls then, too. Mary likes rather loud music and sometimes misses the alarm together. Given the evidence of who has or has not called, we would like to estimate the probability of burglary.
Conditional Distributions

Data Structures:
Conditional Probability Table (CPT)
Each row in a CPT contains the conditional probability of each node value for a conditioning case.

A conditioning case is – a combination of values for the present nodes
(A miniature atomic event)
Property: each row must sum to 1.
That is particularly important for discrete, non-binary variables.
A node without parent has a CPT with only one row: the prior probability of each possible value of the variable.
Semantics of Bayesian Network

Two ways to understanding the semantics of a Bayesian network:

1. See the network as a representation of the joint probability distribution
2. View the Bayesian network as an encoding of a collection of conditional independence statements

They are equivalent. The first view helps in understanding how to construct networks. The second view helps in designing inference procedures.
Representing the full joint distribution

- A Bayesian Network provides a description of the domain. Every entry in the full joint probability distribution ("joint") can be calculated from the information in the network.
- A Generic entry in the joint is the probability of a conjunction of particular assignments to each variable, such as:
  \[ P(X_1 = x_1 \land \ldots \land X_n = x_n) \]  abbreviated as  \[ P(x_1, \ldots, x_n) \]

- The value for the entry is given by
  \[
P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))
\]

  where \( \text{parents}(X_i) \) denotes the specific values of the variables in \( \text{Parents}(X_i) \)

Each configuration  \( (X_1 = x_1 \land \ldots \land X_n = x_n) \) represents a possible world
Each entry in the joint is represented by the product of the appropriate elements of the CPT in the Bayesian Network → the CPTs provide a decomposed representation of the joint.

**Example:** Compute the probability that the alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call.

Denote: $b$ – burglary, $e$ – earthquake, $a$ – alarm has sounded, $j$ – John, $m$ – Mary.

\[
P(j \land m \land a \land \neg b \land \neg e) = P(j|a)P(m|a)P(a|\neg b \land \neg e)P(\neg b)P(\neg e)
\]

\[
= 0.90 \times 0.70 \times 0.00 \times 0.999 \times 0.998
\]

\[
= 0.00062
\]
The Joint for Example 2 (Part 1)

- The Joint Table for variables b, e, a, j, m

<table>
<thead>
<tr>
<th>b</th>
<th>e</th>
<th>a</th>
<th>j</th>
<th>m</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(1-0.001) × (1-0.002) × (1-0.001) × (1-0.05) × (1-0.01)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>(1-0.001) × (1-0.002) × (1-0.001) × (1-0.05) × 0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>(1-0.001) × (1-0.002) × (1-0.001) × 0.05 × (1-0.01)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>(1-0.001) × (1-0.002) × (1-0.001) × 0.05 × 0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>(1-0.001) × (1-0.002) × 0.001 × (1-0.05) × (1-0.01)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>(1-0.001) × (1-0.002) × 0.001 × (1-0.05) × 0.01</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(1-0.001) × (1-0.002) × 0.001 × 0.9 × (1-0.7)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(1-0.001) × (1-0.002) × 0.001 × 0.9 × 0.7</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(1-0.001) × 0.002 × (1-0.29) × (1-0.05) × (1-0.01)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>(1-0.001) × 0.002 × (1-0.29) × (1-0.05) × 0.01</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>(1-0.001) × 0.002 × (1-0.29) × 0.05 × (1-0.01)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>(1-0.001) × 0.002 × (1-0.29) × 0.05 × 0.01</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>(1-0.001) × 0.002 × 0.29 × (1-0.9) × (1-0.7)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>(1-0.001) × 0.002 × 0.29 × (1-0.9) × 0.7</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(1-0.001) × 0.002 × 0.29 × 0.9 × (1-0.7)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(1-0.001) × 0.002 × 0.29 × 0.9 × 0.7</td>
</tr>
</tbody>
</table>
### The Joint for Example 2 (Part 2)

- The Joint Table for variables b, e, a, j, m

<table>
<thead>
<tr>
<th>b</th>
<th>e</th>
<th>a</th>
<th>j</th>
<th>m</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.001 \times (1-0.002) \times (1-0.94) \times (1-0.05) \times (1-0.01)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.001 \times (1-0.002) \times (1-0.94) \times (1-0.05) \times 0.01</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.001 \times (1-0.002) \times (1-0.94) \times 0.05 \times (1-0.01)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.001 \times (1-0.002) \times (1-0.94) \times 0.05 \times 0.01</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.001 \times (1-0.002) \times 0.94 \times (1-0.9) \times (1-0.7)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.001 \times (1-0.002) \times 0.94 \times 0.9 \times 0.7</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.001 \times (1-0.002) \times 0.94 \times 0.9 \times 0.7</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.001 \times (1-0.002) \times 0.94 \times 0.9 \times 0.7</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.001 \times 0.002 \times (1-0.95) \times (1-0.05) \times (1-0.01)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.001 \times 0.002 \times (1-0.95) \times 0.05 \times (1-0.01)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.001 \times 0.002 \times (1-0.95) \times 0.05 \times 0.01</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.001 \times 0.002 \times 0.95 \times (1-0.9) \times (1-0.7)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.001 \times 0.002 \times 0.95 \times (1-0.9) \times 0.7</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.001 \times 0.002 \times 0.95 \times 0.9 \times (1-0.7)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.001 \times 0.002 \times 0.95 \times 0.9 \times 0.7</td>
</tr>
</tbody>
</table>
The Joint for Example 2 (Part 1)

We can compute any probability

\[ P(\neg b, \neg e, \neg a, j, \neg m) = \operatorname{Prob}(0, 1, 0, 1, 0) = (1 - 0.001) \times 0.002 \times (1 - 0.29) \times 0.05 \times (1 - 0.01) \]

\[ P(a | b, \neg e) = \frac{P(a, b, \neg e)}{P(b, \neg e)} = (0.001 \times (1 - 0.002) \times 0.94 \times (1 - 0.9) \times (1 - 0.7) + 0.001 \times (1 - 0.002) \times 0.94 \times (1 - 0.9) \times 0.7 + 0.001 \times (1 - 0.002) \times 0.94 \times (1 - 0.7) \times 0.9 + 0.94 \times (1 - 0.002) \times 0.001 \times 0.7 \times 0.9) / (\text{blue+green}). \]

<table>
<thead>
<tr>
<th>b</th>
<th>e</th>
<th>a</th>
<th>j</th>
<th>m</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(1-0.001)\times(1-0.002)\times(1-0.001)\times(1-0.05)\times(1-0.01) = 0.93674</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>(1-0.001)\times(1-0.002)\times(1-0.001)\times(1-0.05)\times0.01 = 0.00962</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>(1-0.001)\times(1-0.002)\times(1-0.001)\times0.05\times(1-0.01) = 0.04330</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>(1-0.001)\times(1-0.002)\times(1-0.001)\times0.05\times0.01 = 0.000498</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>(1-0.001)\times(1-0.002)\times0.001\times(1-0.05)\times(1-0.01) = 0.000938</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>(1-0.001)\times(1-0.002)\times0.001\times(1-0.05)\times0.01 = 0.00009</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(1-0.001)\times(1-0.002)\times0.001\times0.9\times(1-0.7) = 0.000269</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(1-0.001)\times(1-0.002)\times0.001\times0.9\times0.7 = 0.000281</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(1-0.001)\times0.002\times(1-0.29)\times(1-0.05)\times(1-0.01) = 0.001334</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>(1-0.001)\times0.002\times(1-0.29)\times(1-0.05)\times0.01 = 0.000134</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>(1-0.001)\times0.002\times(1-0.29)\times0.05\times(1-0.01) = 0.0007</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>(1-0.001)\times0.002\times(1-0.29)\times0.05\times0.01 = 0.000070</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>(1-0.001)\times0.002\times0.29\times(1-0.9)\times(1-0.7) = 0.000173</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>(1-0.001)\times0.002\times0.29\times(1-0.9)\times0.7 = 0.000465</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(1-0.001)\times0.002\times0.29\times0.9\times(1-0.7) = 0.0001564</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(1-0.001)\times0.002\times0.29\times0.9\times0.7 = 0.0003649</td>
</tr>
</tbody>
</table>
We are interested in inference that can be computed directly from the network, without the explicit computation of an exponentially long table.

Consider a special network topology: the polytree. A polytree is a directed acyclic graph for which there is just one path along the undirected graph arcs between any two nodes. Any node connected to a node Q will not be connected to any other node except through node Q.

The evidence might include variables associated with components above Q (E⁺) or below Q (E⁻). Then:

\[
\text{Prob}(A₁, …, Aₘ|E⁺) = \text{Prob}(A₁|E⁺) \times \ldots \times \text{Prob}(Aₘ|E⁺) \\
\text{Prob}(B₁, …, Bₖ|E⁻, Q) = \text{Prob}(B₁|E⁻, Q) \times \ldots \times \text{Prob}(Bₖ|E⁻, Q)
\]

Inference is based on Bayes rule for all n possible worlds of variable X:

\[
\text{Prob}(X) = \text{Prob}(X|w₁) \times \text{Prob}(w₁) + \ldots + \text{Prob}(X|wₙ) \times \text{Prob}(wₙ)
\]

There are three cases:

1/ No evidence
2/ Evidence above the query
3/ Evidence below the query
Case 1: No evidence

- A query variable Q is given, and there is no evidence. Goal: compute P(Q)

- Method:
  - 1/ Compute the probability of variables A₁, ..., Aₘ above Q
  - 2/ Compute the probability of all possible worlds wᵢ based on A₁, ..., Aₘ using:

    \[ \text{Prob}(A₁, ..., Aₘ|E⁺) = \text{Prob}(A₁|E⁺) \times \ldots \times \text{Prob}(Aₘ|E⁺) \]

- Compute the probability of Q using:

  \[ \text{Prob}(Q) = \text{Prob}(Q|w₁) \times \text{Prob}(w₁) + \ldots + \text{Prob}(Q|wₙ) \times \text{Prob}(wₙ) \]

Example: Q=B then P(B)=0.001 Q=E then P(E) = 0.002
At node A:

\[
\begin{array}{c|c|c|c}
\hline
B & E & P(A) & P(w) \\
\hline
T & T & 0.95 & 0.001 \times 0.002 = 0.000002 \\
T & F & 0.94 & 0.001 \times 0.998 = 0.000998 \\
F & T & 0.29 & 0.999 \times 0.002 = 0.001998 \\
F & F & 0.001 & 0.999 \times 0.998 = 0.997002 \\
\hline
\end{array}
\]

\[
P(A) = 0.000002 \times 0.95 + 0.000998 \times 0.94 + 0.001998 \times 0.29 + 0.997002 \times 0.001 = 0.002509522
\]
then \( P(\neg A) = 0.99749048 \)
Case 1: No evidence (cont)

Example: \( P(B) = 0.001 \quad P(E) = 0.002 \)

At node A:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.9</td>
<td>0.001 x 0.002 = 0.000002</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>0.001 x 0.998 = 0.000998</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0.999 x 0.002 = 0.001998</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0.999 x 0.998 = 0.997002</td>
</tr>
</tbody>
</table>

\( P(A) = 0.002509522 \) then \( P(\neg A) = 0.99749048 \)

At node J:

<table>
<thead>
<tr>
<th>A</th>
<th>P(J)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.9</td>
<td>0.002509522</td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td>0.99749048</td>
</tr>
</tbody>
</table>

\( P(J) = 0.9 \times 0.002509522 + 0.05 \times 0.99749048 = 0.052133093 \)

\( P(\neg J) = 0.94786091 \)

At node M:

<table>
<thead>
<tr>
<th>A</th>
<th>P(M)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.7</td>
<td>0.002509522</td>
</tr>
<tr>
<td>F</td>
<td>0.01</td>
<td>0.99749048</td>
</tr>
</tbody>
</table>

\( P(M) = 0.7 \times 0.002509522 + 0.01 \times 0.99749048 = 0.01173157 \)

\( P(\neg M) = 0.98826843 \)
Case 2: Evidence above the query

- A query variable Q is given, and
  a) Q is part of evidence \( P(Q|E^+) = 0 \) or \( 1 \) (according to evidence)
  b) Q is at the top, then \( P(Q) \) is given by the CPT
  c) Method:
    - 1/ Compute the probability of variables \( A_1, ..., A_m \) above Q
    - 2/ Compute the probability of all possible worlds \( w_i \) based on \( A_1, ..., A_m \) using:
      \[
      \text{Prob}(A_1, ..., A_m|E^+) = \text{Prob}(A_1|E^+) \times \cdots \text{Prob}(A_m|E^+)
      \]
      - Compute the probability of Q using:
        \[
        \text{Prob}(Q) = \text{Prob}(Q|w_1) \times \text{Prob}(w_1) + \cdots + \text{Prob}(Q|w_n) \times \text{Prob}(w_n)
        \]

Example 1: \( P(A|B) \)
At node A:

\[
\begin{array}{cccc}
<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>0.002 \times 0.95 = 0.002</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>0.001 \times 0.998 = 0.000998</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0 \times 0.002 = 0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0 \times 0.998 = 0</td>
</tr>
</tbody>
</table>
\end{array}
\]

\[
P(A|B) = 0.002 \times 0.95 + 0.998 \times 0.94 = 0.0019 + 0.93812 = 0.94002
\]
then \( P(\neg A|B) = 0.05998 \)
Example 2

Example 2: \( P(M|B) \)

At node A:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>0.002</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>0.998</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0.002</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0.998</td>
</tr>
</tbody>
</table>

\[
P(A|B) = 0.002 \times 0.95 + 0.998 \times 0.94 = 0.0019 + 0.93812 = 0.94002
\]

then \( P(\neg A|B) = 0.05998 \)

At node M:

\[
P(M|B) = 0.7 \times 0.94002 + 0.01 \times 0.05998 = 0.658 + 0.0006 = 0.6586
\]

\[
P(\neg M|B) = 0.3414 \quad \quad P(M|B) = 0.6586
\]
Case 3: Evidence below the query

- A query variable Q is given, and the evidence E^-.
- The goal: compute P(Q|E^-)

Method:
- Compute

\[
P(Q \mid E^-) = \frac{P(E^- \mid Q) \times P(Q)}{P(E^-)}
\]

Thus we need to compute:
1/ P(E^- | Q) – case 2
2/ P(Q) – case 1
3/ a normalization constant

Example 1: P(B|JM)

\[
P(B|JM) = \alpha \times P(J|B) \times P(M|B) \times P(B)
\]

At node A:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>1 × 0.002 = 0.002</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>1 × 0.998 = 0.998</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0 × 0.002 = 0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0 × 0.998 = 0</td>
</tr>
</tbody>
</table>

P(A|B) = 0.002 × 0.95 + 0.998 × 0.94
= 0.0019 + 0.93812 = 0.94002
then P(¬A|B) = 0.05998
Case 3 (Cont)

\[ \text{P(A|B)} = 0.94002 \quad \text{P(¬A|B)} = 0.05998 \]

At node J:

<table>
<thead>
<tr>
<th>A</th>
<th>P(J)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.9</td>
<td>0.94002</td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td>0.05998</td>
</tr>
</tbody>
</table>

\[ \text{P(J|B)} = 0.9 \times 0.94002 + 0.05 \times 0.05998 \]
\[ \approx 0.846 + 0.003 = 0.849 \]
\[ \text{P(¬J|B)} = 0.151 \]
Also P(B)= 0.001

Now \[ \text{P(B|JM)} = \alpha \times \text{P(J|B)} \times \text{P(M|B)} \times \text{P(B)} = \]
\[ = \alpha \times 0.849 \times 0.664 \times 0.001 = \]
\[ \alpha \times 0.00059224 \]

To compute \( \alpha \) we need to compute also

\[ \text{P(¬B|JM)} = \alpha \times \text{P(J|¬B)} \times \text{P(M|¬B)} \times \text{P(¬B)} = \]
\[ = \alpha \times 0.0014919 \]
\[ \alpha(0.00059224 + 0.0014919) = 1; \alpha = 479.8 \]
\[ \text{P(B|JM)} = <0.284, 0.716> \]

\[ \text{P(M|B)} = 0.7 \times 0.94002 + 0.01 \times 0.05998 \]
\[ = 0.658 + 0.0006 = 0.6586 \]
\[ \text{P(¬M|B)} = 0.3414 \]
Exact inference

Basic task: compute the posterior probability distribution for a set of query variables, given some observed event

X → query variable,
E → set of evidence variables $E_1, E_2, \ldots, E_m$

and

e → a particular observed event

Y → non-evidence variables $Y_1, Y_2, \ldots, Y_l$

(also called hidden variables)

The complete set of variables $X = \{x\} \cup E \cup Y$

Typical query: $P(X|e) \leftarrow$ the posterior distribution
Example

Burglary network: event → JohnCalls = true
   MaryCalls = true

Ask if a burglary has occurred:
   \[ P(\text{Burglary} | \text{JohnCalls}=\text{true}, \text{MaryCalls}=\text{true}) = <0.284, 0.716> \]

Inference by enumeration

\[ P(X | e) = \alpha P(X,e) = \alpha \sum_{y} P(X,e,y) \]

For the query \( P(\text{Burglary}|\text{JohnCalls}=\text{true}, \text{MaryCalls}=\text{true}) \)
the hidden variables are Earthquake and Alarm

\[ P(B | j,m) = \alpha P(B, j, e) = \alpha \sum_{e} \sum_{a} P(B, e, a, j, m) \]

The semantics of Bayesian networks gives us:

\[
\begin{align*}
P(b | j,m) &= \alpha \sum_{e} \sum_{a} P(b) P(e) P(a | b, e) P(j | a) P(m | a) \\
\end{align*}
\]
Computing the Posterior

\[ P(b \mid j, m) = \alpha \sum_e \sum_a P(b) P(e) P(a \mid b, c) P(j \mid a) P(m \mid a) \]
\[ = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \]

How do we compute this?

\[ P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \]

At node A:

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>(1 \times 0.002 = 0.000002)</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>(1 \times 0.998 = 0.998)</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0 \times 0.002 = 0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0 \times 0.998 = 0</td>
</tr>
</tbody>
</table>
At node A:

\[ P(b \mid j,m) = \alpha P(b) \sum_{e} P(e) \sum_{a} P(a \mid b,e) P(j \mid a) P(m \mid a) \]

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>P(A)</th>
<th>P(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.95</td>
<td>1 × 0.002 = 0.002</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.94</td>
<td>1 × 0.998 = 0.998</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.29</td>
<td>0 × 0.002 = 0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0 × 0.998 = 0</td>
</tr>
</tbody>
</table>
Final Structure

\[ P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \]
Enumerate-Ask

function ENUMERATION-ASK($X$, $e$, $bn$) returns a distribution over $X$
inputs: $X$, the query variable
$e$, observed values for variables $E$
$bn$, a Bayes net with variables \{$X\} \cup E \cup Y$  /* $Y$ = hidden variables */

$Q(X) \leftarrow$ a distribution over $X$, initially empty
for each value $x_i$ of $X$ do
  extend $e$ with value $x_i$ for $X$
  $Q(x_i) \leftarrow$ ENUMERATE-ALL(VARS[$bn$], $e$)
return normalize($Q(X)$)

function ENUMERATE-ALL($vars$, $e$) returns a real number
if EMPTY?($vars$) then return 1.0
$Y \leftarrow$ FIRST($vars$)
if $Y$ has value $y$ in $e$
  then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $e$)
else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $e_y$)
where $e_y$ is $e$ extended with $Y = y$
Improvements

\[ P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \]

repetitions
Variable Elimination Algorithm

- **Idea**: Eliminate repeated calculations.
- **How?** Do the calculations once and save the results for later use.
- **Evaluate expressions in the right-to-left order (bottom-up in the tree)**

\[
P(B \mid j, m) = \alpha P(B) \sum_{e} P(e) \sum_{a} P(a \mid B, e) P(j \mid a) P(m \mid a)
\]

Factors

- \( f_M(A) = \begin{pmatrix} P(m \mid a) \\ P(m \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix} \)
- \( f_J(A) = \begin{pmatrix} P(j \mid a) \\ P(j \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix} \)
How does one compute the factors???

A set of initial factors:

<table>
<thead>
<tr>
<th>B</th>
<th>f_B</th>
<th>E</th>
<th>f_E</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.001</td>
<td>T</td>
<td>0.002</td>
</tr>
<tr>
<td>F</td>
<td>0.999</td>
<td>F</td>
<td>0.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>f_A</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.95</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>0.94</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>0.06</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>0.29</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>0.71</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0.01</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>J</th>
<th>f_J</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.9</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.1</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.05</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>M</th>
<th>f_M</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.7</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.3</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.01</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.90</td>
</tr>
</tbody>
</table>

For P(B|j,m) we know that J=T and M=T, thus

\[
f_M(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}
\]

\[
f_J(A) = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}
\]
Computing the factors

For $P(a|B,e)$  

\[ f_A(A,B,E) = \]

2×2×2 matrix

From $f_M(A)$, $f_J(A)$ and $f_A(A,B,E)$ we must sum out $A$!

\[ f_{AJM}(B,E) = \sum_a f_A(a,B,E) \times f_J(a) \times f_M(a) = \]

\[ = f_A(a,B,E) \times f_J(a) \times f_M(a) + \]

\[ + f_A(\neg a,B,E) \times f_J(\neg a) \times f_M(\neg a) \]

Use pointwise product

<table>
<thead>
<tr>
<th>B</th>
<th>E</th>
<th>A</th>
<th>$f_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.95</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.05</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>0.94</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>0.06</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>0.29</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>0.71</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0.01</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Pointwise Product

- it is not matrix multiplication!
- it is not element-by-element multiplication!

Pointwise product of two factors $f_1$ and $f_2$ yields a new factor $f$ whose variables are the union of variables $f_1$ and $f_2$. Suppose the two factors have variables $Y_1, \ldots, Y_k$ in common.

Then we have:

$$f(X_1, \ldots, X_j, Y_1, \ldots, Y_k, Z_1, \ldots, Z_e) =$$

$$= f_1(X_1, \ldots, X_j, Y_1, \ldots, Y_k) f_2(Y_1, \ldots, Y_k, Z_1, \ldots, Z_e)$$
Computing Pointwise Product

\[ f(X_1, ..., X_j, Y_1, ..., Y_k, Z_1, ..., Z_e) = f_1(X_1, ..., X_j, Y_1, ..., Y_k) f_2(Y_1, ..., Y_k, Z_1, ..., Z_e) \]

- If all the variables are binary then \( f_1 \) and \( f_2 \) have \( 2^{j+k} \) and \( 2^{k+l} \) entries respectively, and the pointwise product has \( 2^{j+k+l} \) entries.
- Example: Consider \( f_1(A, B) \) and \( f_2(B, C) \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( f_1(A, B) )</th>
<th></th>
<th></th>
<th>( f_2(B, C) )</th>
<th></th>
<th></th>
<th>( f_3(A, B, C) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>.3</td>
<td>T</td>
<td>T</td>
<td>.2</td>
<td>T</td>
<td>T</td>
<td>.3 \times .2</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>.7</td>
<td>T</td>
<td>F</td>
<td>.8</td>
<td>T</td>
<td>T</td>
<td>.3 \times .8</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>.9</td>
<td>F</td>
<td>T</td>
<td>.6</td>
<td>T</td>
<td>F</td>
<td>.7 \times .6</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>.1</td>
<td>F</td>
<td>F</td>
<td>.4</td>
<td>T</td>
<td>F</td>
<td>.7 \times .4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>T</td>
<td>.9 \times .2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>T</td>
<td>.9 \times .8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F</td>
<td>.1 \times .6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F</td>
<td>.1 \times .4</td>
</tr>
</tbody>
</table>

Spring 2016  Dan I. Moldovan, Human Language Technology Research Institute, The University of Texas at Dallas
The Burglary World

\[ f_{A^\uparrow JM}(B, E) = \sum_a f_A(a, B, E) \times f_J(a) \times f_M(a) = \]

\[ = f_A(a, B, E) \times f_J(a) \times f_M(a) + \]

\[ + f_A(\neg a, B, E) \times f_J(\neg a) \times f_M(\neg a) = \]

\[
\begin{array}{c|c|c|c|c}
 B & E & A & f_A \\
\hline
 T & T & T & 0.95 \\
 T & T & F & 0.05 \\
 T & F & T & 0.94 \\
 T & F & F & 0.06 \\
 F & T & T & 0.29 \\
 F & T & F & 0.71 \\
 F & F & T & 0.01 \\
 F & F & F & 0.99 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
 A & J & f_J \\
\hline
 T & T & 0.9 \\
 F & T & 0.05 \\
\end{array}
\]

\[
\begin{array}{c|c|c}
 A & M & f_M \\
\hline
 T & T & 0.7 \\
 F & T & 0.01 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
 B & E & f_{A^\uparrow JM} \\
\hline
 T & T & 0.95 \times 0.7 \times 0.9 + 0.05 \times 0.01 \times 0.05 \\
 T & F & 0.94 \times 0.7 \times 0.9 + 0.06 \times 0.01 \times 0.05 \\
 F & T & 0.29 \times 0.7 \times 0.9 + 0.71 \times 0.01 \times 0.05 \\
 F & F & 0.01 \times 0.7 \times 0.9 + 0.99 \times 0.01 \times 0.05 \\
\end{array}
\]
\[ f_{E\overline{A JM}}(B) = f_E(e) \times f_{\overline{AJM}}(B, e) + f_E(\neg e) \times f_{\overline{AJM}}(B, \neg e) \]

\[
\begin{array}{c|c|c}
B & E & f_{AJM} \\
\hline
T & T & 0.95 \times 0.7 \times 0.9 + 0.05 \times 0.01 \times 0.05 \\
T & F & 0.94 \times 0.7 \times 0.9 + 0.06 \times 0.01 \times 0.05 \\
F & T & 0.29 \times 0.7 \times 0.9 + 0.71 \times 0.01 \times 0.05 \\
F & F & 0.01 \times 0.7 \times 0.9 + 0.99 \times 0.01 \times 0.05 \\
\end{array}
\]
Computing the answer

\[ P(B \mid j, m) = \alpha f_B(B) \times f_{EAJM}(B) \]

<table>
<thead>
<tr>
<th>B</th>
<th>f_B</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.001</td>
</tr>
<tr>
<td>F</td>
<td>0.999</td>
</tr>
</tbody>
</table>

\[ 0.002 \times (0.95 \times 0.7 \times 0.9 + 0.05 \times 0.01 \times 0.05) + 0.998 \times (0.94 \times 0.7 \times 0.9 + 0.06 \times 0.01 \times 0.05) \]

\[ 0.002 \times (0.29 \times 0.7 \times 0.9 + 0.71 \times 0.01 \times 0.05) + 0.998 \times (0.01 \times 0.7 \times 0.9 + 0.99 \times 0.01 \times 0.05) \]
Elimination-Ask

```plaintext
function Elimination-Ask(X, e, bn) returns a distribution over X
    inputs: X, the query variable
            e, evidence specified as an event
            bn, a Bayesian network specifying joint distribution P(X1, ..., Xn)

    factors ← []; vars ← REVERSE(VARS[bn])
    for each var in vars do
        factors ← [MAKE-FACTOR(var, e)|factors]
        if var is a hidden variable then factors ← SUM-OUT(var, factors)
    return NORMALIZE(POINTWISE-PRODUCT(factors))
```
A Method for Constructing Bayesian Networks

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]

defines what a given Bayesian Network means

- It does not explain how to build a Bayesian Network such that the resulting joint distribution is a good representation of a given domain
- However, it implies certain additional independence relationships that can be used to guide the knowledge engineer in constructing the topology of the network
Steps

1. Rewrite the joint distribution in terms of a conditional probability using the product rule:

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]

becomes

\[ P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \]

2. Repeat the process, reducing each conjunctive probability to a conditional probability and a smaller conjunction →

\[ P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots P(x_2 \mid x_1) P(x_1) \]

\[ = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1) \]

(the chain rule)
Semantics

Because

\[ P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_1) \ldots P(x_2 \mid x_1) P(x_1) \]

\[ = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1) \]

We have:

\[ P(x_i \mid x_{i-1}, \ldots, x_1) = P(x_i \mid \text{Parents}(x_i)) \]

if \( \text{Parents}(x_i) \subseteq \{x_{i-1}, x_1\} \)

Nodes must be labeled in any order that is consistent with the partial order implicit in the graph structure
Construction Rules

- The parents of node $X_i$ should contain all nodes $X_1, \ldots, X_{i-1}$ that directly influence $X_i$
- The correct order in which to add nodes: add “root causes” first, then variables they influence, and so on, until reaching the leaves that have no direct causal influence on other variables
- What happens if we chose the wrong order of adding nodes?
Problematic Networks

If we add the nodes in the order: MaryCalls, JohnCalls, Alarm, Burglary, Earthquake we obtain a more complicated network:

The process:
1) Add MaryCalls – no parents
2) Add JohnCalls → if MaryCalls, probably the alarm went off → it makes it more likely JohnCalls → JohnCalls needs MaryCalls as a parent
3) Add Alarm → if both John and Mary call, it is more likely the alarm went off → both MaryCalls and JohnCalls are parents
More Problems

4) Add Burglary → if we know the alarm state, then the call from John or Mary might give us information about our phone ringing or Mary’s music, but not about the burglary →

\[ P(\text{Burglary} \mid \text{Alarm}, \text{JohnCalls, MaryCalls}) = P(\text{Burglary}, \text{Alarm}) \]

→ only Alarm is a parent

5) Adding Earthquake: if the alarm is on, it is more likely that there has been an earthquake. But if we know that there has been a burglary, then that explains the alarm. Both Alarm and Burglary are parents
What is the Problem?

→ We do not have only causal links!
→ There are also links from symptoms to Causes!
Conditional Independence in Bayesian Networks

- A node is conditionally independent of its predecessors, given its parents because:

\[ P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i)) \]

- Conditional independence is also dictated by topological semantics:
  1. A node is conditionally independent of its non-descendants, given its parents. Example: JohnCalls is independent of Burglary and Earthquake, given the value of Alarm.
  2. A node is conditionally independent of all other nodes in the network, given its parents, children and children’s parents \( \rightarrow \text{called Markov blanket} \).
Examples

(a) A node X is conditionally independent of its non-descendants (e.g., the Z_{ij}s) given its parents (the U_i s shown in the gray area).

(b) A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area)
Efficient Representation of Conditional Distribution

- Even when a node has only $K$ parents $\rightarrow$ the CPT needs $O(2^K)$ members $\rightarrow$ this is the worst-case scenario. Usually, relations between parents and child are described by a canonical distribution that fits some standard pattern $\rightarrow$ the CPT can be specified by naming the pattern and few parameters.

- Example: deterministic nodes. A deterministic node has its value specified exactly by the value of its parents, with no uncertainty.
  - Example:

```
Canadian       US       Mexican
    \downarrow   \downarrow   \downarrow
North-American
```

(a disjunction)
Noisy-OR

- Uncertain relationship – characterized by “noisy” logical relationships. Noisy-OR relation: a generalization of the logical OR

- Example
  - In propositional logic: Fever = Cold ∨ Flu ∨ Malaria
  - Noisy-OR: allows for uncertainty about the ability of each parent to cause the child to be true.
  - The relation between parent ant child may be inhibited e.q. Cold = true but Fever = false

- The inhibition of each parent is independent of the inhibition of any other parent

- All possible causes are listed
### Example:

**Probabilities of individual inhibitions**

\[
P(\neg \text{fever} | \neg \text{cold}, \neg \text{flu}, \neg \text{malaria}) = 0.6
\]
\[
P(\neg \text{fever} | \neg \text{cold}, \text{flu}, \neg \text{malaria}) = 0.2
\]
\[
P(\neg \text{fever} | \neg \text{cold}, \neg \text{flu}, \text{malaria}) = 0.1
\]

From this information, the entire CPT can be built

<table>
<thead>
<tr>
<th>Cold</th>
<th>Flu</th>
<th>Malaria</th>
<th>P(Fever)</th>
<th>P(\neg \text{Fever})</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>0.98</td>
<td>0.02 = 0.2 x 0.1</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>0.94</td>
<td>0.06 = 0.6 x 0.1</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>0.88</td>
<td>0.12 = 0.6 x 0.2</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>0.988</td>
<td>0.012 = 0.6 x 0.2 x 0.1</td>
</tr>
</tbody>
</table>
Clustering Algorithms

- If we want to compute posterior probabilities for all variables in the network – the variable elimination algorithm is not efficient.
- Solution: Consider clustering algorithms also known as join tree algorithms.
- The idea: Cluster individual modes of the network to form cluster nodes ⇒ the resulting network is a poly tree.