Inference in Bayesian networks

Jiří Vomlel

Academy of Sciences of the Czech Republic

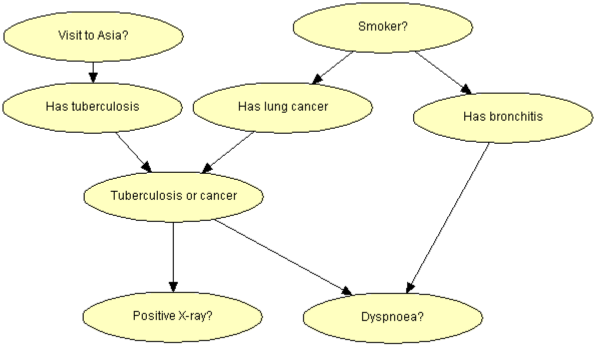
9th July, 2007

Simple diagnostic example - 1

A medical doctor inspects a patient.
Possible diagnosis are: tuberculosis, lung cancer, or bronchitis.

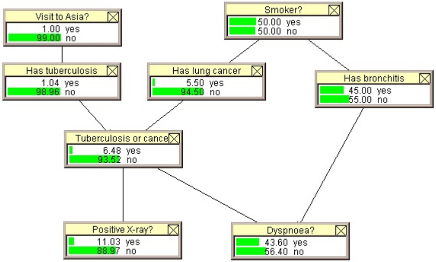
Simple diagnostic example - 1

A medical doctor inspects a patient.
Possible diagnosis are: tuberculosis, lung cancer, or bronchitis.



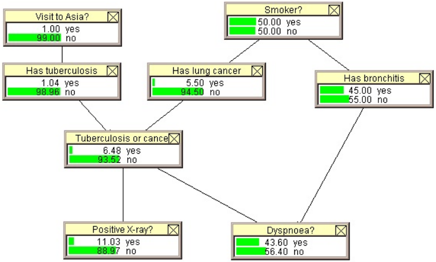
Simple diagnostic example - 2

We do not know anything about the patient.

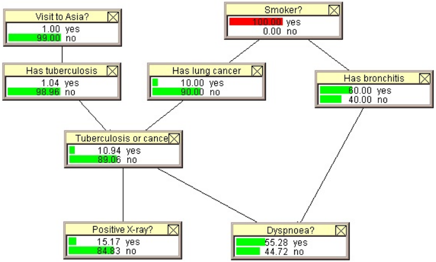


Simple diagnostic example - 2

We do not know anything about the patient.

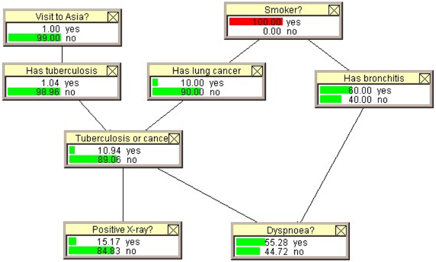


The patient is a smoker.



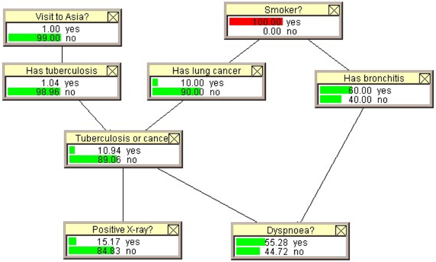
Simple diagnostic example - 3

The patient is a smoker.

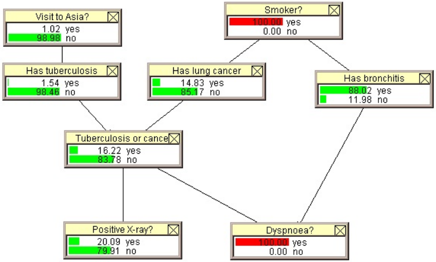


Simple diagnostic example - 3

The patient is a smoker.

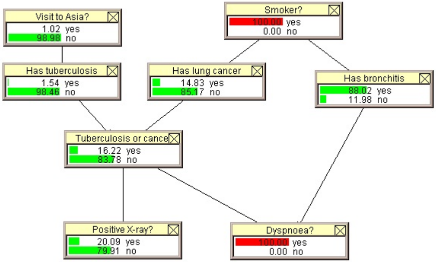


... and he suffers from dyspnoea.



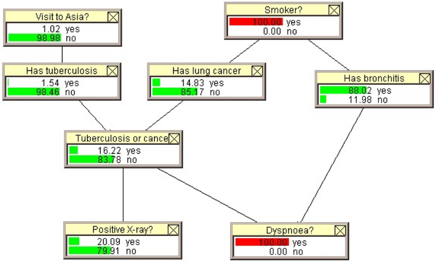
Simple diagnostic example - 4

The patient is a smoker
and he suffers from dyspnoea.

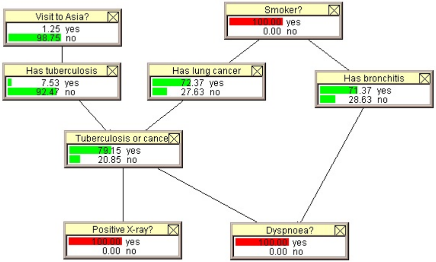


Simple diagnostic example - 4

The patient is a smoker
and he suffers from dyspnoea.

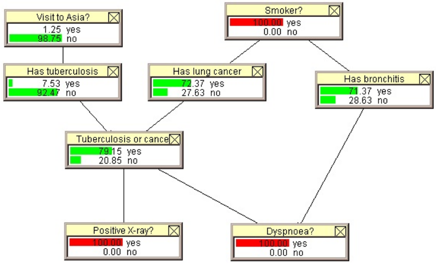


... and his X-ray is positive



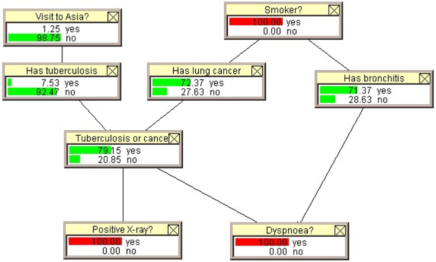
Simple diagnostic example - 5

The patient is a smoker,
he suffers from dyspnoea
and his X-ray is positive

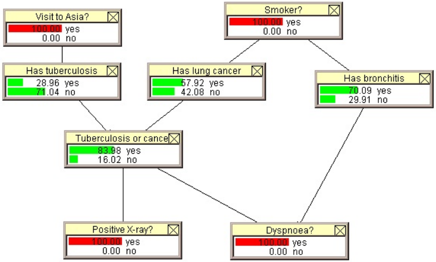


Simple diagnostic example - 5

The patient is a smoker,
he suffers from dyspnoea
and his X-ray is positive

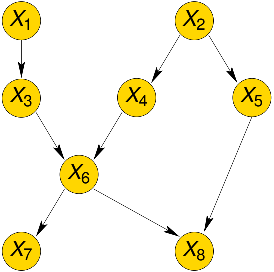


... and he visited Asia recently.



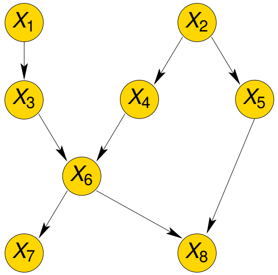
Example written more formally

X_1	"Visit to Asia?"
X_2	"Smoker?"
X_3	"Has tuberculosis?"
X_4	"Has lung cancer?"
X_5	"Has bronchitis?"
X_6	"Tuberculosis or cancer"
X_7	"Positive X-ray?"
X_8	"Dyspnoea"



Example written more formally

X_1	"Visit to Asia?"
X_2	"Smoker?"
X_3	"Has tuberculosis?"
X_4	"Has lung cancer?"
X_5	"Has bronchitis?"
X_6	"Tuberculosis or cancer"
X_7	"Positive X-ray?"
X_8	"Dyspnoea"

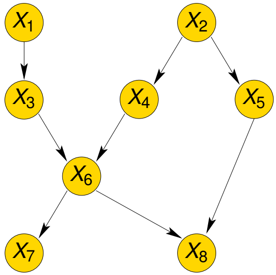


The joint probability distribution defined by the Bayesian network:

$$P(X_1, X_2, \ldots, X_8) \quad = \quad \prod_{i=1}^{8} P(X_i | \{X_j\}_{j \in Pa(i)})$$

Example written more formally

X_1	"Visit to Asia?"
X_2	"Smoker?"
X_3	"Has tuberculosis?"
X_4	"Has lung cancer?"
X_5	"Has bronchitis?"
X_6	"Tuberculosis or cancer"
X_7	"Positive X-ray?"
X_8	"Dyspnoea"



The joint probability distribution defined by the Bayesian network:

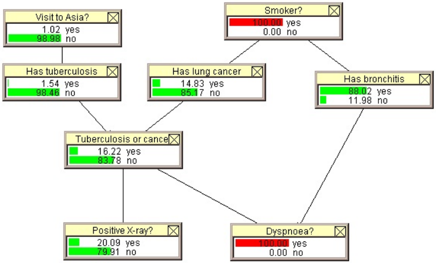
$$
\begin{aligned}
P(X_1, X_2, \ldots, X_8) &= \prod_{i=1}^{8} P(X_i | \{X_j\}_{j \in Pa(i)}) \\
&= P(X_8|X_6, X_5) \cdot P(X_7|X_6) \cdot P(X_6|X_3, X_4) \\
&\quad \cdot P(X_5|X_2) \cdot P(X_4|X_2) \cdot P(X_3|X_1) \cdot P(X_2) \cdot P(X_1)
\end{aligned}
$$

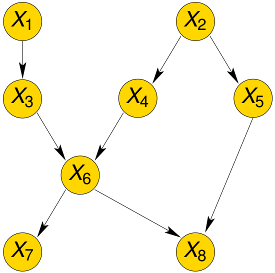
Conditional probability

“What is the probability that the patient has tuberculosis given he is a smoker and suffers from dyspnoea?”

Conditional probability

"What is the probability that the patient has tuberculosis given he is a smoker and suffers from dyspnoea?"



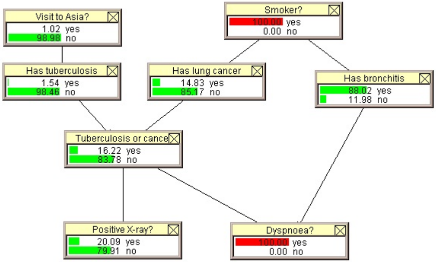


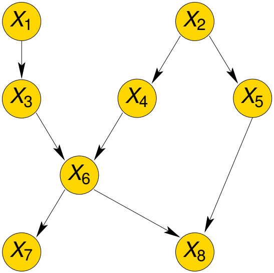
The conditional probability distribution corresponding to the query:

$$P(X_3, | X_2 = 1, X_8 = 1) \quad = \quad \frac{P(X_2 = 1, X_3, X_8 = 1)}{P(X_2 = 1, X_8 = 1)}$$

Conditional probability

“What is the probability that the patient has tuberculosis given he is a smoker and suffers from dyspnoea?”



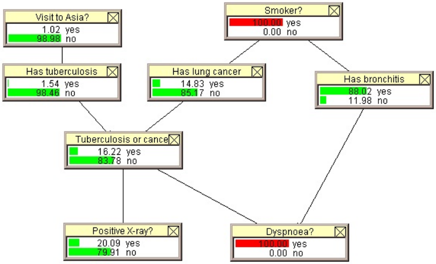


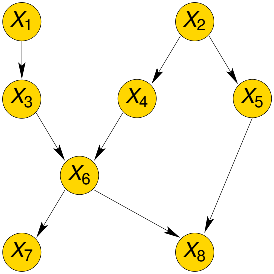
The conditional probability distribution corresponding to the query:

$$P(X_3, |X_2 = 1, X_8 = 1) \quad = \quad \frac{P(X_2 = 1, X_3, X_8 = 1)}{\sum_{X_3} P(X_2 = 1, X_3, X_8 = 1)}$$

Conditional probability

"What is the probability that the patient has tuberculosis given he is a smoker and suffers from dyspnoea?"





The conditional probability distribution corresponding to the query:

$$P(X_3, | X_2 = 1, X_8 = 1) = \frac{P(X_2 = 1, X_3, X_8 = 1)}{\sum_{X_3} P(X_2 = 1, X_3, X_8 = 1)}$$

$$P(X_2 = 1, X_3, X_8 = 1) = \sum_{X_1, X_4, X_5, X_6, X_7} P(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1)$$

Distributive law for probability distributions

$$P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline\end{array} \cdot \begin{array}{|c|cc|}\hline & Z=0 & Z=1 \\ \hline Y=0 & e & g \\ Y=1 & f & h \\ \hline\end{array}$$

Distributive law for probability distributions

$$P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|} \hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline \end{array} \cdot \begin{array}{|c|cc|} \hline & Z=0 & Z=1 \\ \hline Y=0 & e & g \\ Y=1 & f & h \\ \hline \end{array}$$

$$= \begin{array}{|c|cccc|} \hline & \multicolumn{2}{c}{X=0} & \multicolumn{2}{c|}{X=1} \\ & Z=0 & Z=1 & Z=0 & Z=1 \\ \hline Y=0 & ae & ag & ce & cg \\ Y=1 & bf & bh & df & dh \\ \hline \end{array}$$

Distributive law for probability distributions

$$P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline \end{array} \cdot \begin{array}{|c|cc|}\hline & Z=0 & Z=1 \\ \hline Y=0 & e & g \\ Y=1 & f & h \\ \hline \end{array}$$

$$= \begin{array}{|c|cccc|}\hline & \multicolumn{2}{c}{X=0} & \multicolumn{2}{c|}{X=1} \\ & Z=0 & Z=1 & Z=0 & Z=1 \\ \hline Y=0 & ae & ag & ce & cg \\ Y=1 & bf & bh & df & dh \\ \hline \end{array}$$

$$\sum_Z P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & ae+ag & ce+cg \\ Y=1 & bf+bh & df+dh \\ \hline \end{array}$$

Distributive law for probability distributions

$$P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline\end{array} \cdot \begin{array}{|c|cc|}\hline & Z=0 & Z=1 \\ \hline Y=0 & e & g \\ Y=1 & f & h \\ \hline\end{array}$$

$$= \begin{array}{|c|cccc|}\hline & \multicolumn{2}{c}{X=0} & \multicolumn{2}{c|}{X=1} \\ & Z=0 & Z=1 & Z=0 & Z=1 \\ \hline Y=0 & ae & ag & ce & cg \\ Y=1 & bf & bh & df & dh \\ \hline\end{array}$$

$$\sum_Z P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & ae+ag & ce+cg \\ Y=1 & bf+bh & df+dh \\ \hline\end{array}$$

$$= \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline\end{array} \cdot \begin{array}{|c|c|}\hline Y=0 & e+g \\ Y=1 & f+h \\ \hline\end{array}$$

Distributive law for probability distributions

$$P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline\end{array} \cdot \begin{array}{|c|cc|}\hline & Z=0 & Z=1 \\ \hline Y=0 & e & g \\ Y=1 & f & h \\ \hline\end{array}$$

$$= \begin{array}{|c|cccc|}\hline & X=0 & & X=1 & \\ & Z=0 & Z=1 & Z=0 & Z=1 \\ \hline Y=0 & ae & ag & ce & cg \\ Y=1 & bf & bh & df & dh \\ \hline\end{array}$$

$$\sum_Z P(X,Y)\cdot P(Y,Z) = \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & ae+ag & ce+cg \\ Y=1 & bf+bh & df+dh \\ \hline\end{array}$$

$$= \begin{array}{|c|cc|}\hline & X=0 & X=1 \\ \hline Y=0 & a & c \\ Y=1 & b & d \\ \hline\end{array} \cdot \begin{array}{|c|c|}\hline Y=0 & e+g \\ Y=1 & f+h \\ \hline\end{array}$$

$$= P(X,Y)\cdot\left(\sum_Z P(Y,Z)\right)$$

Direct computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1)$$
$$= \sum_{X_1, X_4, X_5, X_6, X_7} P(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1)$$

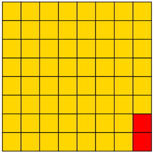
Direct computation of a marginal probability

$$
\begin{aligned}
& P(X_2 = 1, X_3, X_8 = 1) \\
& = \sum_{X_1, X_4, X_5, X_6, X_7} P(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1) \\
& = \sum_{X_1, X_4, X_5, X_6, X_7} \left(\begin{array}{l} P(X_8 = 1 | X_6, X_5) \cdot P(X_7 | X_6) \cdot P(X_6 | X_3, X_4) \\ \cdot P(X_5 | X_2 = 1) \cdot P(X_4 | X_2 = 1) \cdot P(X_3 | X_1) \\ \cdot P(X_2 = 1) \cdot P(X_1) \end{array} \right)
\end{aligned}
$$

Direct computation of a marginal probability

$$
\begin{aligned}
& P(X_2 = 1, X_3, X_8 = 1) \\
& = \sum_{X_1,X_4,X_5,X_6,X_7} P(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1) \\
& = \sum_{X_1,X_4,X_5,X_6,X_7} \left(\begin{array}{l} P(X_8 = 1 | X_6, X_5) \cdot P(X_7 | X_6) \cdot P(X_6 | X_3, X_4) \\ \cdot P(X_5 | X_2 = 1) \cdot P(X_4 | X_2 = 1) \cdot P(X_3 | X_1) \\ \cdot P(X_2 = 1) \cdot P(X_1) \end{array} \right)
\end{aligned}
$$

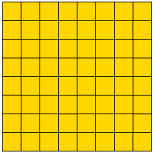
- $\psi(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1)$
  $\rightarrow \psi(X_2 = 1, X_3, X_8 = 1)$



Direct computation of a marginal probability

$$
\begin{aligned}
& P(X_2 = 1, X_3, X_8 = 1) \\
& \quad = \sum_{X_1, X_4, X_5, X_6, X_7} P(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1) \\
& \quad = \sum_{X_1, X_4, X_5, X_6, X_7} \left(\begin{array}{l} P(X_8 = 1 | X_6, X_5) \cdot P(X_7 | X_6) \cdot P(X_6 | X_3, X_4) \\ \cdot P(X_5 | X_2 = 1) \cdot P(X_4 | X_2 = 1) \cdot P(X_3 | X_1) \\ \cdot P(X_2 = 1) \cdot P(X_1) \end{array} \right)
\end{aligned}
$$

- $\psi(X_1, X_2 = 1, X_3, \ldots, X_7, X_8 = 1)$
  $\rightarrow \psi(X_2 = 1, X_3, X_8 = 1)$



The largest table has size $2^6 = 64$.

Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

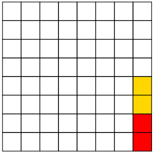
$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3|X_1) \cdot P(X_1) \right)$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} \left(P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1) \right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1) \right) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \to \psi(X_6)$

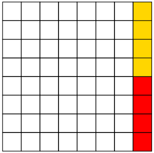


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \rightarrow \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \rightarrow \psi(X_3, X_6)$

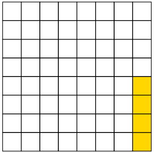


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3|X_1) \cdot P(X_1) \right)$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} {\color{red}\sum_{X_7} P(X_7|X_6)} \\ \cdot \sum_{X_4} \left(P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1) \right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1) \right) \end{array} \right)$$

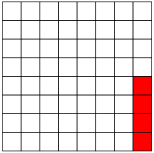
- $\sum_{X_5} \psi(X_5, X_6) \to \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \to \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \to 1$



Efficient computation of a marginal probability

$$
\begin{aligned}
P(X_2 = 1, X_3, X_8 = 1) \quad &= P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3 | X_1) \cdot P(X_1) \right) \\
&\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7 | X_6) \\ \cdot \sum_{X_4} \left(P(X_6 | X_3, X_4) \cdot P(X_4 | X_2 = 1) \right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1 | X_6, X_5) \cdot P(X_5 | X_2 = 1) \right) \end{array} \right)
\end{aligned}
$$

- $\sum_{X_5} \psi(X_5, X_6) \rightarrow \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \rightarrow \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \rightarrow 1$
- $\psi(X_6) \cdot \psi(X_3, X_6) \rightarrow \psi'(X_3, X_6)$

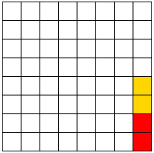


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \rightarrow \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \rightarrow \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \rightarrow 1$
- $\psi(X_6) \cdot \psi(X_3, X_6) \rightarrow \psi'(X_3, X_6)$
- $\sum_{X_6} \psi'(X_3, X_6) \rightarrow \psi(X_3)$

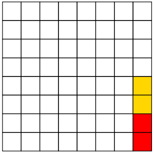


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3 | X_1) \cdot P(X_1)\right)$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7 | X_6) \\ \cdot \sum_{X_4} \left(P(X_6 | X_3, X_4) \cdot P(X_4 | X_2 = 1)\right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1 | X_6, X_5) \cdot P(X_5 | X_2 = 1)\right) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \rightarrow \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \rightarrow \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \rightarrow 1$
- $\psi(X_6) \cdot \psi(X_3, X_6) \rightarrow \psi'(X_3, X_6)$
- $\sum_{X_6} \psi'(X_3, X_6) \rightarrow \psi(X_3)$
- $\sum_{X_1} \psi(X_1, X_3) \rightarrow \psi'(X_3)$

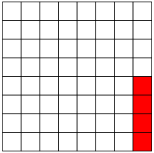


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \rightarrow \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \rightarrow \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \rightarrow 1$
- $\psi(X_6) \cdot \psi(X_3, X_6) \rightarrow \psi'(X_3, X_6)$
- $\sum_{X_6} \psi'(X_3, X_6) \rightarrow \psi(X_3)$
- $\sum_{X_1} \psi(X_1, X_3) \rightarrow \psi'(X_3)$
- $\psi(X_3) \cdot \psi'(X_3) \cdot P(X_2 = 1) \rightarrow P(X_2 = 1, X_3, X_8 = 1)$

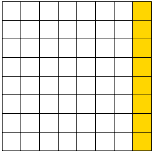


Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

- $\sum_{X_5} \psi(X_5, X_6) \to \psi(X_6)$
- $\sum_{X_4} \psi(X_3, X_4, X_6) \to \psi(X_3, X_6)$
- $\sum_{X_7} \psi(X_6, X_7) \to 1$
- $\psi(X_6) \cdot \psi(X_3, X_6) \to \psi'(X_3, X_6)$
- $\sum_{X_6} \psi'(X_3, X_6) \to \psi(X_3)$
- $\sum_{X_1} \psi(X_1, X_3) \to \psi'(X_3)$
- $\psi(X_3) \cdot \psi'(X_3) \cdot P(X_2 = 1) \to P(X_2 = 1, X_3, X_8 = 1)$

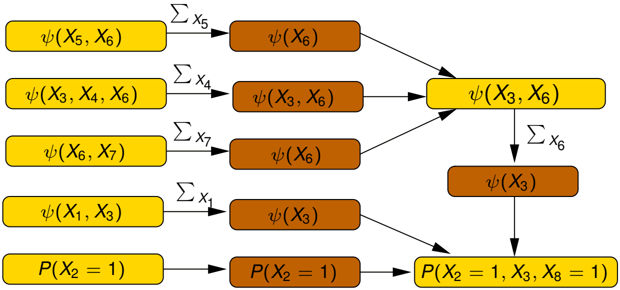


The largest table has size $2^3 = 8$.

Efficient computation of a marginal probability

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$



Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?
- What shall we do if we want to compute all marginals?

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?
- What shall we do if we want to compute all marginals?

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?
- What shall we do if we want to compute all marginals?

Looking closer:

- Generally, computing a marginal probability in a Bayesian network is NP-hard.

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?
- What shall we do if we want to compute all marginals?

Looking closer:

- Generally, computing a marginal probability in a Bayesian network is NP-hard.
- However, for some families of Bayesian network models we have efficient procedures.

Efficient computation of marginal probabilities

- How can we find an ordering of summations and products that requires only small tables during computation?
- Can we reuse the ordering if we want to compute other marginals?
- What shall we do if we want to compute all marginals?

Looking closer:

- Generally, computing a marginal probability in a Bayesian network is NP-hard.
- However, for some families of Bayesian network models we have efficient procedures.
- Junction Tree Algorithm

Domain graph

Recall, the joint probability distribution of the Bayesian network

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \{X_j\}_{j \in Pa(i)})$$

Domain graph

Recall, the joint probability distribution of the Bayesian network

$$P(X_1, \ldots, X_n) \quad = \quad \prod_{i=1}^{n} P(X_i | \{X_j\}_{j \in Pa(i)})$$

$D_i = \{X_i\} \cup \{X_j\}_{j \in Pa(i)}$ is the domain of $P(X_i | \{X_j\}_{j \in Pa(i)})$.

Domain graph

Recall, the joint probability distribution of the Bayesian network

$$P(X_1, \ldots, X_n) \quad = \quad \prod_{i=1}^{n} P(X_i | \{X_j\}_{j \in Pa(i)})$$

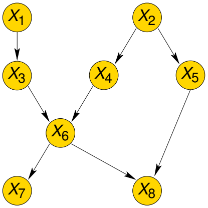
$D_i = \{X_i\} \cup \{X_j\}_{j \in Pa(i)}$ is the domain of $P(X_i | \{X_j\}_{j \in Pa(i)})$.
The domain graph is an undirected graph with variables $X_1, \ldots, X_n$ as nodes and with an edge between a pair of variables X_a, X_b if there exists $P(X_i | \{X_j\}_{j \in Pa(i)})$ such that $X_a, X_b \in D_i$.

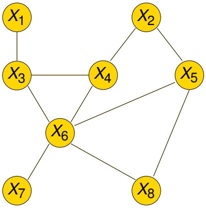
Domain graph

Recall, the joint probability distribution of the Bayesian network

$$P(X_1, \ldots, X_n) \;=\; \prod_{i=1}^{n} P(X_i | \{X_j\}_{j \in Pa(i)})$$

$D_i = \{X_i\} \cup \{X_j\}_{j \in Pa(i)}$ is the domain of $P(X_i | \{X_j\}_{j \in Pa(i)})$.
The domain graph is an undirected graph with variables $X_1, \ldots, X_n$ as nodes and with an edge between a pair of variables X_a, X_b if there exists $P(X_i | \{X_j\}_{j \in Pa(i)})$ such that $X_a, X_b \in D_i$.



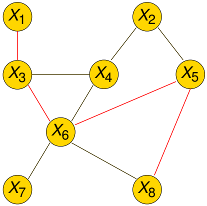


Triangulated graph

A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.

Triangulated graph

A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.

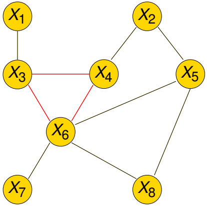


Triangulated graph

A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.
A cycle is a sequence $A_1, \ldots, A_{\ell+1} = A_1$ of nodes, where $A_1, \ldots, A_\ell$ is a path and $\{A_\ell, A_{\ell+1}\}$ is an edge; ℓ is the length of the cycle.

Triangulated graph

A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.
A cycle is a sequence $A_1, \ldots, A_{\ell+1} = A_1$ of nodes, where $A_1, \ldots, A_\ell$ is a path and $\{A_\ell, A_{\ell+1}\}$ is an edge; ℓ is the length of the cycle.

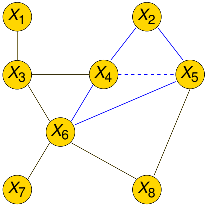


Triangulated graph

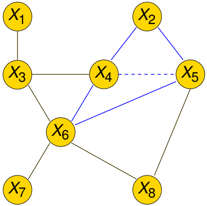
A path is a sequence $A_1, \dots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.
A cycle is a sequence $A_1, \dots, A_{\ell+1} = A_1$ of nodes, where $A_1, \dots, A_\ell$ is a path and $\{A_\ell, A_{\ell+1}\}$ is an edge; ℓ is the length of the cycle.
A chord in a cycle $A_1, \dots, A_{\ell+1}$ is an edge between two not consecutive nodes in the cycle.

Triangulated graph

A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.
A cycle is a sequence $A_1, \ldots, A_{\ell+1} = A_1$ of nodes, where $A_1, \ldots, A_\ell$ is a path and $\{A_\ell, A_{\ell+1}\}$ is an edge; ℓ is the length of the cycle.
A chord in a cycle $A_1, \ldots, A_{\ell+1}$ is an edge between two not consecutive nodes in the cycle.



A path is a sequence $A_1, \ldots, A_\ell$ of distinct nodes connected by edges $\{A_i, A_{i+1}\}$.
A cycle is a sequence $A_1, \ldots, A_{\ell+1} = A_1$ of nodes, where $A_1, \ldots, A_\ell$ is a path and $\{A_\ell, A_{\ell+1}\}$ is an edge; ℓ is the length of the cycle.
A chord in a cycle $A_1, \ldots, A_{\ell+1}$ is an edge between two not consecutive nodes in the cycle.



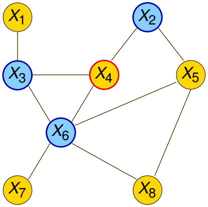
An undirected graph is triangulated if it does not contain a cycle of length four or more without a chord.

Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.

Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.

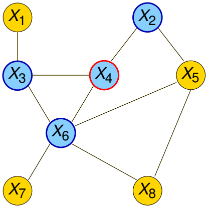


Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.

Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.



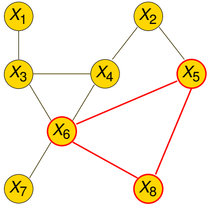
Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.

Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.

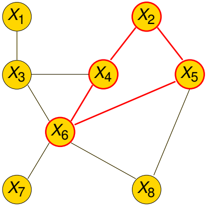
a complete set



Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.

not a complete set

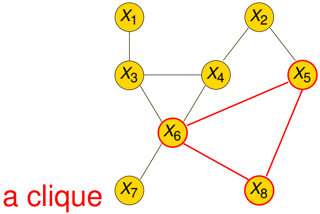


Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.

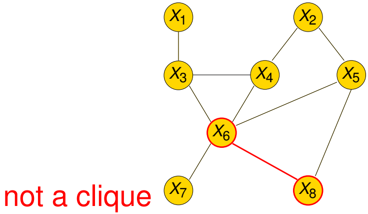
Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.



Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.

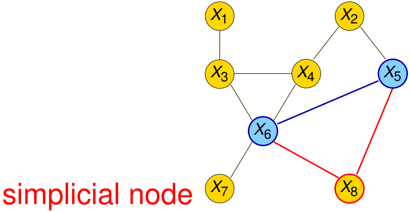


Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.
A node with a complete neighbor set is simplicial.

Graph-theoretic concepts

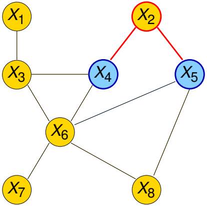
Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.
A node with a complete neighbor set is simplicial.



Graph-theoretic concepts

Nodes connected by an edge to node X are called neighbors of X.
Neighbors of X plus X are family of X.
A set of nodes is complete if all nodes are pairwise connected by an edge.
A maximal complete set with respect to set inclusion is called clique.
A node with a complete neighbor set is simplicial.

not a simplicial node



Elimination sequence

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

Elimination sequence

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

Elimination sequence

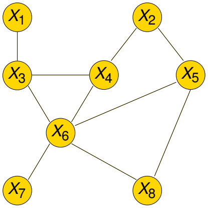
$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

domain graph



Elimination sequence

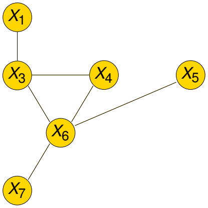
$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

nodes with evidence eliminated



Elimination sequence

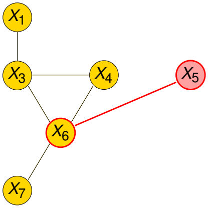
$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

eliminating X_5



Elimination sequence

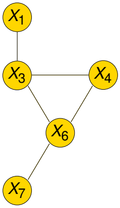
$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

X_5 eliminated



Elimination sequence

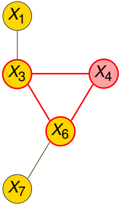
$$P(X_2=1,X_3,X_8=1) \;=\; P(X_2=1)\cdot\sum_{X_1}\left(P(X_3|X_1)\cdot P(X_1)\right)$$

$$\cdot\sum_{X_6}\left(\begin{array}{l}\sum_{X_7}P(X_7|X_6)\\ \cdot\sum_{X_4}\left(P(X_6|X_3,X_4)\cdot P(X_4|X_2=1)\right)\\ \cdot\sum_{X_5}\left(P(X_8=1|X_6,X_5)\cdot P(X_5|X_2=1)\right)\end{array}\right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

eliminating X_4



Elimination sequence

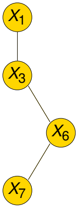
$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

X_4 eliminated



Elimination sequence

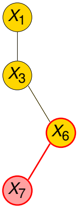
$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

eliminating X_7



Elimination sequence

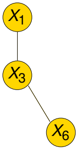
$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

X_7 eliminated



Elimination sequence

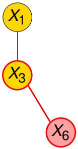
$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

eliminating X_6



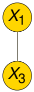
Elimination sequence

$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3 | X_1) \cdot P(X_1) \right)$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} \left(P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1) \right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1) \right) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$



X_6 eliminated

Elimination sequence

$$P(X_2 = 1, X_3, X_8 = 1) \quad = P(X_2 = 1) \cdot \sum_{X_1} (P(X_3|X_1) \cdot P(X_1))$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} (P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)) \\ \cdot \sum_{X_5} (P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

eliminating X_1

Elimination sequence

$$P(X_2 = 1, X_3, X_8 = 1) \;\; = P(X_2 = 1) \cdot \sum_{X_1} \left(P(X_3|X_1) \cdot P(X_1)\right)$$

$$\cdot \sum_{X_6} \left(\begin{array}{l} \sum_{X_7} P(X_7|X_6) \\ \cdot \sum_{X_4} \left(P(X_6|X_3, X_4) \cdot P(X_4|X_2 = 1)\right) \\ \cdot \sum_{X_5} \left(P(X_8 = 1|X_6, X_5) \cdot P(X_5|X_2 = 1)\right) \end{array} \right)$$

An elimination sequence of the above computation was

$$X_5, X_4, X_7, X_6, X_1$$

X_1 eliminated

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.
- The added edges are called fill-ins.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.
- The added edges are called fill-ins.
- The more fill-ins we have the larger tables we get during the computation.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.
- The added edges are called fill-ins.
- The more fill-ins we have the larger tables we get during the computation.
- Elimination sequence of variables is better (requires less space) if it introduces less fill-ins.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.
- The added edges are called fill-ins.
- The more fill-ins we have the larger tables we get during the computation.
- Elimination sequence of variables is better (requires less space) if it introduces less fill-ins.
- Perfect elimination sequence is elimination sequence that does not introduce any fill-ins.

Perfect elimination sequence

- When eliminating a variable X we need to multiply all probability tables containing this variable.
- The domain of this product is the family of X.
- When eliminating X we pairwise connect all neighbors of X.
- The added edges are called fill-ins.
- The more fill-ins we have the larger tables we get during the computation.
- Elimination sequence of variables is better (requires less space) if it introduces less fill-ins.
- Perfect elimination sequence is elimination sequence that does not introduce any fill-ins.
- The sequence of the previous example was perfect.

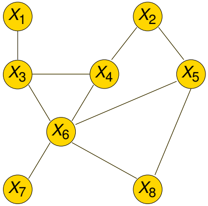
A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

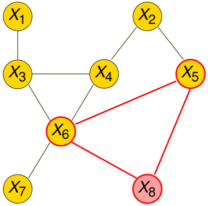
domain graph



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

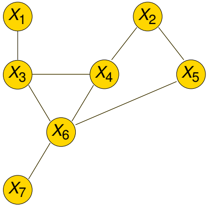
eliminating X_8



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

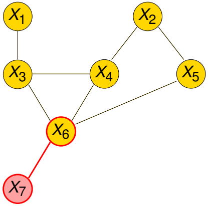
X_8 eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

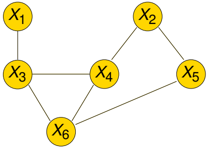
eliminating X_7



A non-perfect sequence

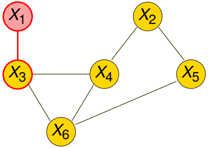
Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

X_7 eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

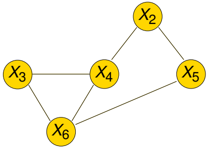


eliminating X_1

A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

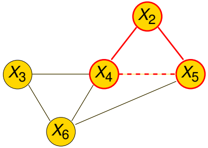
X_1 eliminated
now, there is no simplicial node among nodes to be eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

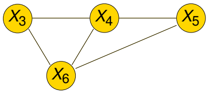
eliminating X_2
filling in edge $\{X_4, X_5\}$



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

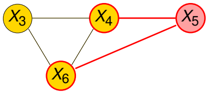
X_2 eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

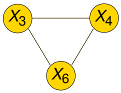
eliminating X_5



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

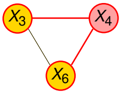
X_5 eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

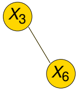
eliminating X_4



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

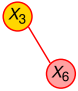
X_4 eliminated



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

eliminating X_6



A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

X_6 eliminated

A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

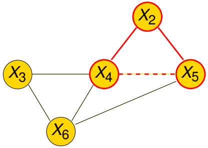
Observe that the added edge is exactly the same as the one we added to make the graph triangulated.

A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

Observe that the added edge is exactly the same as the one we added to make the graph triangulated.

eliminating X_2
filling in edge $\{X_4, X_5\}$

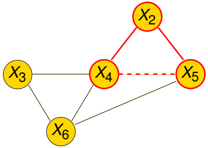


A non-perfect sequence

Compute $P(X_3)$, i.e., a marginal probability of X_3 without any evidence.

Observe that the added edge is exactly the same as the one we added to make the graph triangulated.

eliminating X_2
filling in edge $\{X_4, X_5\}$



In a triangulated graph we can **always** find a perfect elimination sequence!

Join tree

Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph

Join tree

Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)

Join tree

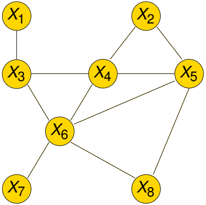
Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)
- for all pairs of nodes C_1 and C_2 it holds that all cliques corresponding to nodes on the path between C_1 and C_2 contain $C_1 \cap C_2$.

Join tree

Join tree of an undirected graph is a undirected graph such that:

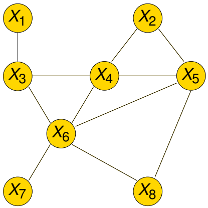
- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)
- for all pairs of nodes C_1 and C_2 it holds that all cliques corresponding to nodes on the path between C_1 and C_2 contain $C_1 \cap C_2$.

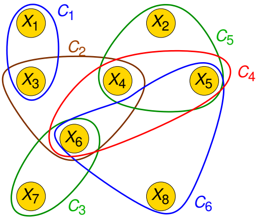


Join tree

Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)
- for all pairs of nodes C_1 and C_2 it holds that all cliques corresponding to nodes on the path between C_1 and C_2 contain $C_1 \cap C_2$.

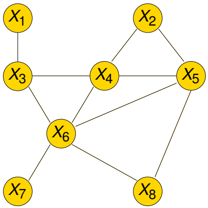


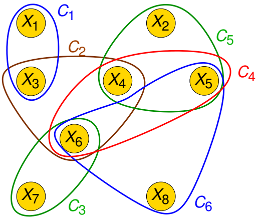


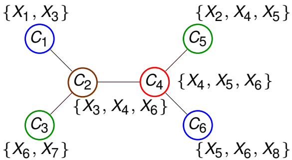
Join tree

Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)
- for all pairs of nodes C_1 and C_2 it holds that all cliques corresponding to nodes on the path between C_1 and C_2 contain $C_1 \cap C_2$.



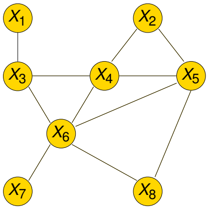


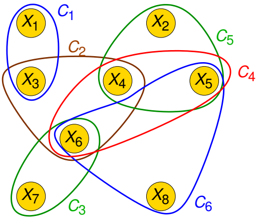


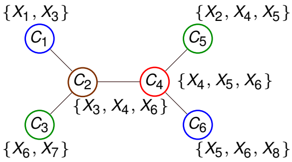
Join tree

Join tree of an undirected graph is a undirected graph such that:

- nodes corresponds to cliques of the graph
- nodes are connected by edges so that the graph is a tree (connected graph without cycles)
- for all pairs of nodes C_1 and C_2 it holds that all cliques corresponding to nodes on the path between C_1 and C_2 contain $C_1 \cap C_2$.







For a triangulated graph we can **always** construct a join tree.

Junction tree

Junction tree is a join tree extended by:

- probability tables attached to a node whose clique contains its domain

Junction tree

Junction tree is a join tree extended by:

- probability tables attached to a node whose clique contains its domain
- to each edge $\{C_i, C_j\}$ the separator set $C_i \cap C_j$ is attached

Junction tree

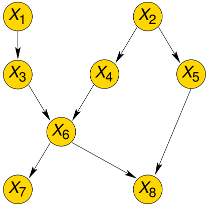
Junction tree is a join tree extended by:

- probability tables attached to a node whose clique contains its domain
- to each edge $\{C_i, C_j\}$ the separator set $C_i \cap C_j$ is attached
- each separator contains two mailboxes for probability tables - for each direction one

Junction tree

Junction tree is a join tree extended by:

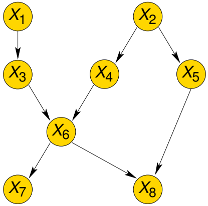
- probability tables attached to a node whose clique contains its domain
- to each edge $\{C_i, C_j\}$ the separator set $C_i \cap C_j$ is attached
- each separator contains two mailboxes for probability tables - for each direction one

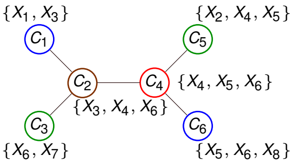


Junction tree

Junction tree is a join tree extended by:

- probability tables attached to a node whose clique contains its domain
- to each edge $\{C_i, C_j\}$ the separator set $C_i \cap C_j$ is attached
- each separator contains two mailboxes for probability tables - for each direction one

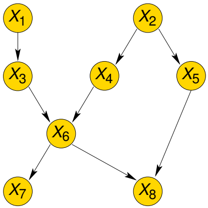


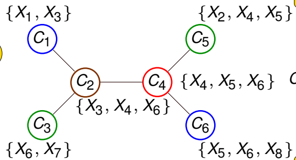


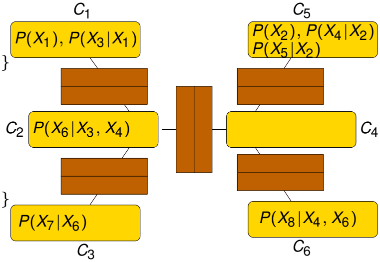
Junction tree

Junction tree is a join tree extended by:

- probability tables attached to a node whose clique contains its domain
- to each edge $\{C_i, C_j\}$ the separator set $C_i \cap C_j$ is attached
- each separator contains two mailboxes for probability tables - for each direction one







Computation of marginal probabilities in junction tree

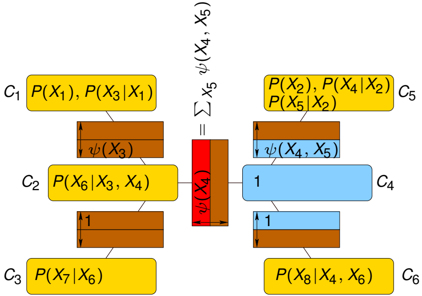
The rules for sending messages:

- each node can send message only if it has received messages in all mailboxes except the mailbox where it is going to send the message,

Computation of marginal probabilities in junction tree

The rules for sending messages:

- each node can send message only if it has received messages in all mailboxes except the mailbox where it is going to send the message,



Computation of marginal probabilities in junction tree

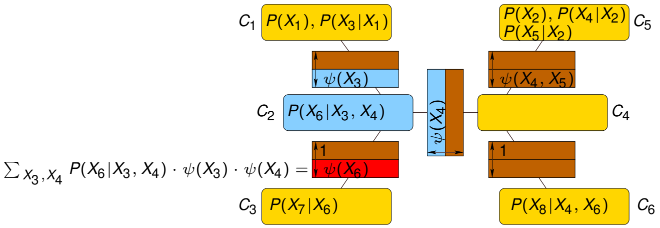
The rules for sending messages:

- each node can send message only if it has received messages in all mailboxes except the mailbox where it is going to send the message,
- a message from C_i to C_j is computed by marginalizing to $C_i \cap C_j$ the product of all probability tables attached to C_i and all tables in incoming mailboxes of C_i except the mailbox from C_j to C_i .

Computation of marginal probabilities in junction tree

The rules for sending messages:

- each node can send message only if it has received messages in all mailboxes except the mailbox where it is going to send the message,
- a message from C_i to C_j is computed by marginalizing to $C_i \cap C_j$ the product of all probability tables attached to C_i and all tables in incoming mailboxes of C_i except the mailbox from C_j to C_i .



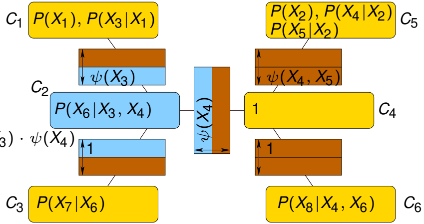
Computation of marginal probabilities in junction tree

- To get a marginal probability table we just need to:
  (1) find a clique containing its domain,
  (2) compute the product of all probability tables attached to this clique and all tables in incoming mailboxes, and
  (3) marginalize it to the domain of the marginal, if necessary.

Computation of marginal probabilities in junction tree

- To get a marginal probability table we just need to:
  (1) find a clique containing its domain,
  (2) compute the product of all probability tables attached to this clique and all tables in incoming mailboxes, and
  (3) marginalize it to the domain of the marginal, if necessary.

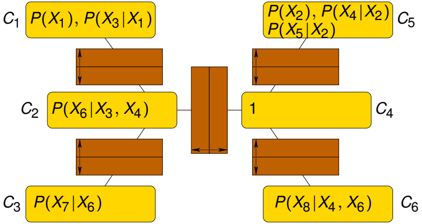
$$P(X_3) = \sum_{X_4, X_6} P(X_6|X_3, X_4) \cdot \psi(X_3) \cdot \psi(X_4)$$



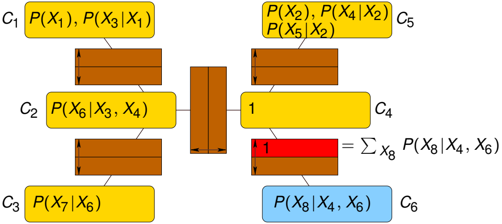
Computation of marginal probabilities in junction tree

- To get a marginal probability table we just need to:
  (1) find a clique containing its domain,
  (2) compute the product of all probability tables attached to this clique and all tables in incoming mailboxes, and
  (3) marginalize it to the domain of the marginal, if necessary.
- The process is finished when all nodes have all mailboxes full.

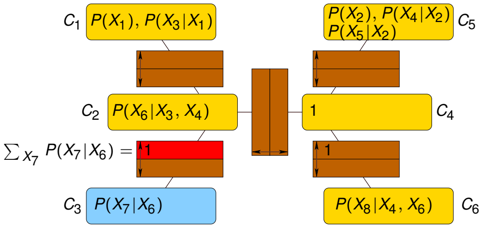
Example of the computation of $P(X_3)$ in junction tree



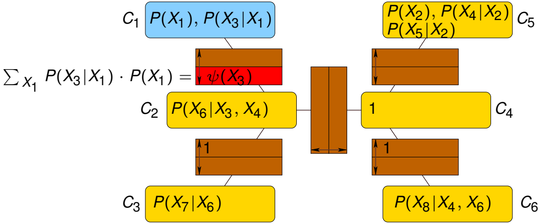
Example of the computation of $P(X_3)$ in junction tree



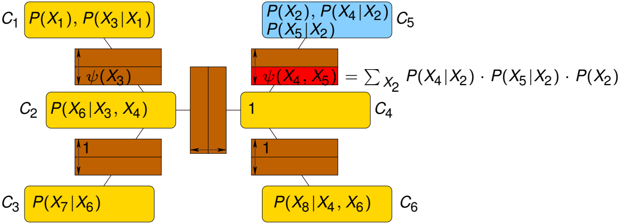
Example of the computation of $P(X_3)$ in junction tree



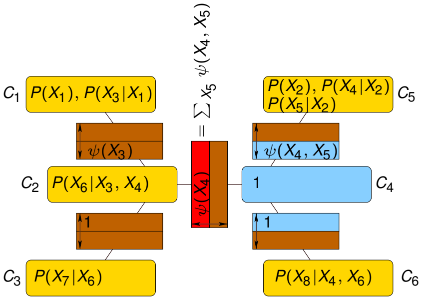
Example of the computation of $P(X_3)$ in junction tree



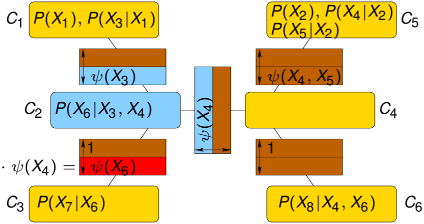
Example of the computation of $P(X_3)$ in junction tree



Example of the computation of $P(X_3)$ in junction tree



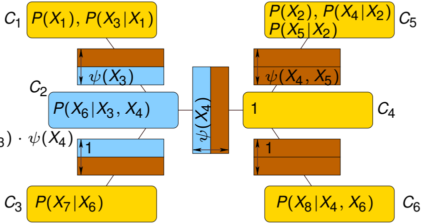
Example of the computation of $P(X_3)$ in junction tree



$\sum_{X_3, X_4} P(X_6|X_3, X_4) \cdot \psi(X_3) \cdot \psi(X_4) =$

Example of the computation of $P(X_3)$ in junction tree

$$P(X_3) = \sum_{X_4, X_6} P(X_6|X_3, X_4) \cdot \psi(X_3) \cdot \psi(X_4)$$



Other inference methods

- Exact methods: Lauritzen-Spiegelhalter method, Shenoy-Shafer method, Lazy propagation (A. Madsen and F. V. Jensen), Variable elimination - e.g., Bucket elimination (R. Dechter et al.).

Other inference methods

- Exact methods: Lauritzen-Spiegelhalter method, Shenoy-Shafer method, Lazy propagation (A. Madsen and F. V. Jensen), Variable elimination - e.g., Bucket elimination (R. Dechter et al.).
- Approximate methods: Markov Chain Monte Carlo (MCMC) methods - e.g., Gibbs sampling, Penniless propagation (A. Cano, S. Moral, and A. Salmerón), Loopy belief propagation (J. Pearl), Variational methods (T. Jaakkola and M. Jordan).

Other inference methods

- Exact methods: Lauritzen-Spiegelhalter method, Shenoy-Shafer method, Lazy propagation (A. Madsen and F. V. Jensen), Variable elimination - e.g., Bucket elimination (R. Dechter et al.).
- Approximate methods: Markov Chain Monte Carlo (MCMC) methods - e.g., Gibbs sampling, Penniless propagation (A. Cano, S. Moral, and A. Salmerón), Loopy belief propagation (J. Pearl), Variational methods (T. Jaakkola and M. Jordan).
- Exact methods exploiting local structure of tables: Algebraic circuits (A. Darwiche et al.).

Other tasks solved in Bayesian networks

- Maximum aposteriori configuration (MAP)

Other tasks solved in Bayesian networks

- Maximum aposteriori configuration (MAP)
- Sensitivity analysis

Other tasks solved in Bayesian networks

- Maximum aposteriori configuration (MAP)
- Sensitivity analysis
- Decision making maximizing expected utility when also utility and decision nodes are included (decision diagrams)

References

Textbook:

- Finn V. Jensen, Bayesian networks and Decision Graphs. Springer-Verlag, 2001.

References

Textbook:

- Finn V. Jensen, Bayesian networks and Decision Graphs. Springer-Verlag, 2001.
- Finn V. Jensen and Thomas D. Nielsen, Bayesian networks and Decision Graphs. Springer-Verlag, Second edition, 2007. (see `http://bndg.cs.aau.dk`)

References

Textbook:

- Finn V. Jensen, Bayesian networks and Decision Graphs. Springer-Verlag, 2001.
- Finn V. Jensen and Thomas D. Nielsen, Bayesian networks and Decision Graphs. Springer-Verlag, Second edition, 2007. (see `http://bndg.cs.aau.dk`)

Software:

References

Textbook:

- Finn V. Jensen, Bayesian networks and Decision Graphs. Springer-Verlag, 2001.
- Finn V. Jensen and Thomas D. Nielsen, Bayesian networks and Decision Graphs. Springer-Verlag, Second edition, 2007. (see `http://bndg.cs.aau.dk`)

Software:

- Hugin. Comercial, free demo version (size limited), `http://www.hugin.com`

References

Textbook:

- Finn V. Jensen, Bayesian networks and Decision Graphs. Springer-Verlag, 2001.
- Finn V. Jensen and Thomas D. Nielsen, Bayesian networks and Decision Graphs. Springer-Verlag, Second edition, 2007. (see `http://bndg.cs.aau.dk`)

Software:

- Hugin. Comercial, free demo version (size limited), `http://www.hugin.com`
- Genie. Freely available, `http://genie.sis.pitt.edu/`