Rank of tensors of l-out-of-k functions: an application in probabilistic inference

Jiří Vomlel

Institute of Information Theory and Automation (ÚTIA)
Academy of Sciences of the Czech Republic

• The computer game of Minesweeper

- The computer game of Minesweeper
- Probabilistic reasoning given evidence (using a simple example)

- The computer game of Minesweeper
- Probabilistic reasoning given evidence (using a simple example)
- Improving the computational efficiency

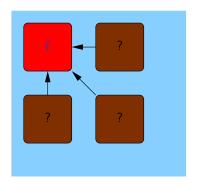
- The computer game of Minesweeper
- Probabilistic reasoning given evidence (using a simple example)
- Improving the computational efficiency
- Rank-one decomposition of probability tables representing addition

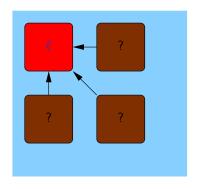
- The computer game of Minesweeper
- Probabilistic reasoning given evidence (using a simple example)
- Improving the computational efficiency
- Rank-one decomposition of probability tables representing addition
- Results of experiments

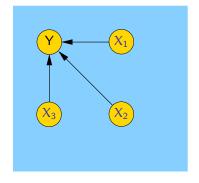
The game of Minesweeper

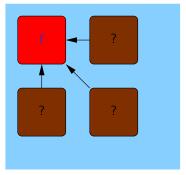




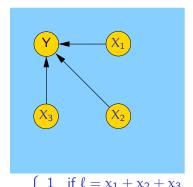


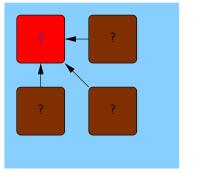






$$P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) = \begin{cases} 1 & \text{if } \ell = x_1 + x_2 + x_3 \\ 0 & \text{otherwise.} \end{cases}$$





$$=\begin{cases} 1 & \text{if } \ell = x_1 + x_2 + x_3 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \frac{r}{}$$

$$P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) = \begin{cases} 1 & \text{if } \ell = x_1 + x_2 + x_3 \\ 0 & \text{otherwise.} \end{cases}$$

$$P(X_i) = \frac{r}{s \cdot t - o}$$

r is the number of mines, o is the number of observations s, t are the dimensions of the game grid.

• Assume we observe $Y = \ell$.

- Assume we observe $Y = \ell$.
- We compute by Bayes rule

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | Y = \ell)$$

- Assume we observe $Y = \ell$.
- We compute by Bayes rule

$$\begin{split} P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | Y = \ell) \\ = & \frac{P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \cdot \prod_{i=1}^3 P(X_i = x_i)}{P(Y = \ell)} \end{split}$$

- Assume we observe $Y = \ell$.
- We compute by Bayes rule

$$\begin{split} P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | Y = \ell) \\ &= \frac{P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \cdot \prod_{i=1}^{3} P(X_i = x_i)}{P(Y = \ell)} \\ &\propto P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \end{split}$$

- Assume we observe $Y = \ell$.
- We compute by Bayes rule

$$\begin{split} P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | Y = \ell) \\ = & \frac{P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \cdot \prod_{i=1}^3 P(X_i = x_i)}{P(Y = \ell)} \\ \propto & P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \end{split}$$

This is a probability table over 3 binary variables X₁, X₂, X₃:

$$\begin{split} P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ = \begin{cases} 1 & \text{if } x_1 + x_2 + x_3 = \ell \\ 0 & \text{otherwise.} \end{cases} \end{split}$$

- Assume we observe $Y = \ell$.
- We compute by Bayes rule

$$\begin{split} P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | Y = \ell) \\ = & \frac{P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \cdot \prod_{i=1}^3 P(X_i = x_i)}{P(Y = \ell)} \\ \propto & P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \end{split}$$

• This is a probability table over 3 binary variables X_1, X_2, X_3 :

$$\begin{split} P(Y = \ell | X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ = & \left\{ \begin{array}{l} 1 \quad \text{if } x_1 + x_2 + x_3 = \ell \\ 0 \quad \text{otherwise.} \end{array} \right. \\ = & \left. \psi(X_1 = x_1, X_2 = x_2, X_3 = x_3) \right. \end{split}$$

Tensors of ℓ-out-of-k functions

We can visualize probability table ψ as a tensor (for $\ell = 1$):

$$\left(\begin{array}{cc} \left(\begin{array}{cc} 0\\1\end{array}\right) & \left(\begin{array}{cc} 1\\0\end{array}\right)\\ \left(\begin{array}{cc} 1\\0\end{array}\right) & \left(\begin{array}{cc} 0\\0\end{array}\right) \end{array}\right)$$

In this talk all tensors are functions from $\{0,1\}^k$ to real numbers.

Tensors of ℓ -out-of-k functions

We can visualize probability table ψ as a tensor (for $\ell = 1$):

$$\left(\begin{array}{cc} \left(\begin{array}{cc} 0\\1\end{array}\right) & \left(\begin{array}{cc} 1\\0\end{array}\right)\\ \left(\begin{array}{cc} 1\\0\end{array}\right) & \left(\begin{array}{cc} 0\\0\end{array}\right) \end{array}\right)$$

In this talk all tensors are functions from $\{0,1\}^k$ to real numbers. We are interested in tensors of ℓ -out-of-k functions $f_\ell(x_1,\ldots,x_k)$, where:

- ℓ is the observed state of Y and
- k is the number of binary variables parents of Y.

Tensors of ℓ -out-of-k functions

We can visualize probability table ψ as a tensor (for $\ell = 1$):

$$\left(\begin{array}{cc} \left(\begin{array}{cc} 0\\1\end{array}\right) & \left(\begin{array}{cc} 1\\0\end{array}\right)\\ \left(\begin{array}{cc} 1\\0\end{array}\right) & \left(\begin{array}{cc} 0\\0\end{array}\right) \end{array}\right)$$

In this talk all tensors are functions from $\{0,1\}^k$ to real numbers. We are interested in tensors of ℓ -out-of-k functions $f_{\ell}(x_1,\ldots,x_k)$, where:

- k is the number of binary variables parents of Y.

$$f_{\ell}(x_1,\ldots,x_k) \ = \ \left\{ \begin{array}{ll} 1 & \text{if } \ell = \sum_{i=1}^k x_i \\ 0 & \text{otherwise}. \end{array} \right.$$

Tensors of ℓ-out-of-k functions

We can visualize probability table ψ as a tensor (for $\ell = 1$):

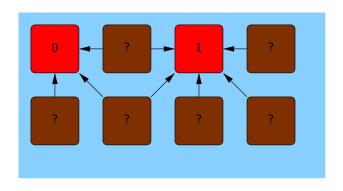
$$\left(\begin{array}{cc} \left(\begin{array}{cc} 0\\1\end{array}\right) & \left(\begin{array}{cc} 1\\0\end{array}\right)\\ \left(\begin{array}{cc} 1\\0\end{array}\right) & \left(\begin{array}{cc} 0\\0\end{array}\right) \end{array}\right)$$

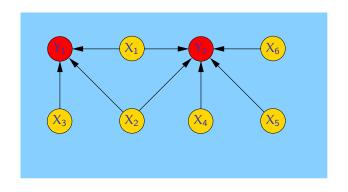
In this talk all tensors are functions from $\{0,1\}^k$ to real numbers. We are interested in tensors of ℓ -out-of-k functions $f_{\ell}(x_1,\ldots,x_k)$, where:

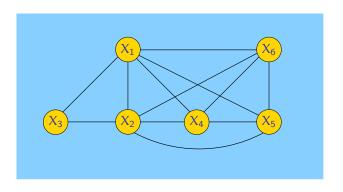
- ℓ is the observed state of Y and
- k is the number of binary variables parents of Y.

$$f_{\ell}(x_1, \dots, x_k) = \begin{cases} 1 & \text{if } \ell = \sum_{i=1}^k x_i \\ 0 & \text{otherwise.} \end{cases}$$

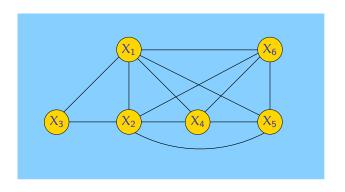
In our example $\ell = 1$ and k = 3.







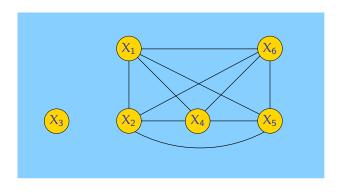
$$\xi(X_1,\dots,X_6) \ = \ \psi(X_1,\dots,X_3) \cdot \phi(X_1,X_2,X_4,\dots,X_6)$$



$$\xi(X_1,\dots,X_6) \ = \ \psi(X_1,\dots,X_3) \cdot \phi(X_1,X_2,X_4,\dots,X_6)$$

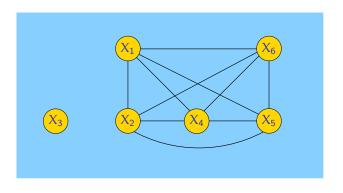
Total table size is $2^3 + 2^5 = 8 + 32 = 40$.

A more efficient way of combining information



$$\begin{array}{rcl} \xi(X_1, \dots, X_6) & = & \psi_1(X_1) \cdot \dots \cdot \psi_3(X_3) \\ & & \cdot \phi_1(X_1, X_2, X_4, \dots, X_6) \end{array}$$

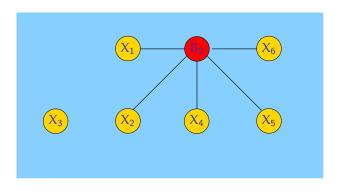
A more efficient way of combining information



$$\begin{array}{rcl} \xi(X_1, \dots, X_6) & = & \psi_1(X_1) \cdot \dots \cdot \psi_3(X_3) \\ & & \cdot \phi_1(X_1, X_2, X_4, \dots, X_6) \end{array}$$

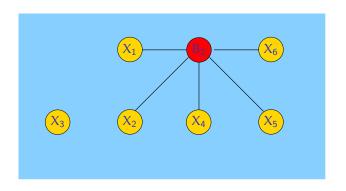
Total table size is $3 \cdot 2 + 2^5 = 6 + 32 = 38$.

An even more efficient way of combining information



$$\begin{split} \xi(X_1,\dots,X_6) &= \sum_{B_2} \psi_1(X_1) \cdot \dots \cdot \psi_3(X_3) \\ &\cdot \phi_1(B_2,X_1) \cdot \phi_2(B_2,X_2) \cdot \phi_4(B_2,X_4) \dots \phi_6(B_2,X_6) \end{split}$$

An even more efficient way of combining information



$$\begin{array}{lll} \xi(X_1,\ldots,X_6) & = & \displaystyle \sum_{B_2} \psi_1(X_1) \cdot \ldots \cdot \psi_3(X_3) \\ \\ & \cdot \phi_1(B_2,X_1) \cdot \phi_2(B_2,X_2) \cdot \phi_4(B_2,X_4) \ldots \phi_6(B_2,X_6) \end{array}$$

Since B is binary the total table size is $3 \cdot 2 + 5 \cdot 2^2 = 6 + 20 = 26$.

We have just seen that

$$\begin{split} \phi_1(X_1, X_2, X_4, \dots, X_6) \\ &= \sum_{B_2} \phi_1(B_2, X_1) \cdot \phi_2(B_2, X_2) \cdot \phi_4(B_2, X_4) \dots \phi_6(B_2, X_6) \ . \end{split}$$

We have just seen that

$$\begin{split} \phi_1(X_1,X_2,X_4,\dots,X_6) \\ &= \sum_{B_2} \phi_1(B_2,X_1) \cdot \phi_2(B_2,X_2) \cdot \phi_4(B_2,X_4) \dots \phi_6(B_2,X_6) \ . \end{split}$$

But there is no way we can write

$$\phi_1(X_1, X_2, X_4, \dots, X_6) \ = \ \phi_1(X_1) \cdot \phi_2(X_2) \cdot \phi_4(X_4) \dots \phi_6(X_6)$$

We have just seen that

$$\begin{split} \phi_1(X_1,X_2,X_4,\dots,X_6) \\ &= \sum_{B_2} \phi_1(B_2,X_1) \cdot \phi_2(B_2,X_2) \cdot \phi_4(B_2,X_4) \dots \phi_6(B_2,X_6) \ . \end{split}$$

But there is no way we can write

$$\phi_1(X_1, X_2, X_4, \dots, X_6) \ = \ \phi_1(X_1) \cdot \phi_2(X_2) \cdot \phi_4(X_4) \dots \phi_6(X_6)$$

What is the minimal number of states of a variable \boldsymbol{B} so that it holds that

$$\psi(X_1,...,X_k) = \sum_{B} \prod_{i=1}^{k} \psi_i(B,X_i) ?$$

We have just seen that

$$\begin{split} \phi_1(X_1,X_2,X_4,\dots,X_6) \\ &= \sum_{B_2} \phi_1(B_2,X_1) \cdot \phi_2(B_2,X_2) \cdot \phi_4(B_2,X_4) \dots \phi_6(B_2,X_6) \ . \end{split}$$

But there is no way we can write

$$\phi_1(X_1, X_2, X_4, \dots, X_6) \ = \ \phi_1(X_1) \cdot \phi_2(X_2) \cdot \phi_4(X_4) \dots \phi_6(X_6)$$

What is the minimal number of states of a variable \boldsymbol{B} so that it holds that

$$\psi(X_1,...,X_k) = \sum_{B} \prod_{i=1}^{k} \psi_i(B,X_i) ?$$

This number is called the rank of tensor ψ .

Symmetric rank of tensors of ℓ -out-of-k functions

• Generally, finding the rank of a tensor is NP-hard.

Symmetric rank of tensors of ℓ -out-of-k functions

- Generally, finding the rank of a tensor is NP-hard.
- However, tensors of ℓ-out-of-k functions define a restricted class of tensors.

Symmetric rank of tensors of ℓ -out-of-k functions

- Generally, finding the rank of a tensor is NP-hard.
- However, tensors of ℓ-out-of-k functions define a restricted class of tensors.
- These tensors are all symmetric. A tensor ψ is symmetric if $\psi(X_1=x_1,\ldots,X_k=x_k)=a_{x_1+\ldots+x_k}$ where $a=(a_0,\ldots,a_k)$ is a vector of real numbers.

Symmetric rank of tensors of ℓ -out-of-k functions

- Generally, finding the rank of a tensor is NP-hard.
- However, tensors of ℓ-out-of-k functions define a restricted class of tensors.
- These tensors are all symmetric. A tensor ψ is symmetric if $\psi(X_1=x_1,\ldots,X_k=x_k)=a_{x_1+\ldots+x_k}$ where $a=(a_0,\ldots,a_k)$ is a vector of real numbers.
- The symmetric rank of tensor ψ is the minimum number of symmetric tensors of rank one that sum up to ψ .

Symmetric rank of tensors of ℓ -out-of-k functions

- Generally, finding the rank of a tensor is NP-hard.
- However, tensors of ℓ-out-of-k functions define a restricted class of tensors.
- These tensors are all symmetric. A tensor ψ is symmetric if $\psi(X_1 = x_1, \dots, X_k = x_k) = a_{x_1 + \dots + x_k}$ where $a = (a_0, \dots, a_k)$ is a vector of real numbers.
- The symmetric rank of tensor ψ is the minimum number of symmetric tensors of rank one that sum up to ψ .

Theorem

The symmetric rank of a tensor representing an ℓ -out-of-k function (for $0 < \ell < k$) is at least $\max\{\ell+1, k-\ell\}$.

Border rank of tensors of ℓ -out-of-k functions

Definition (Border rank)

The border rank of a tensor A is

$$\text{min}\{r: \forall \epsilon>0 \ \exists \text{ tensor } E: \|E\|<\epsilon, \text{rank}(\boldsymbol{A}+\boldsymbol{E})=r\}$$
 ,

where $\|\cdot\|$ is any norm.

Border rank of tensors of ℓ-out-of-k functions

Definition (Border rank)

The border rank of a tensor A is

$$\text{min}\{r: \forall \epsilon>0 \ \exists \text{ tensor } E: \|E\|<\epsilon, \text{rank}(\boldsymbol{A}+\boldsymbol{E})=r\}$$
 ,

where $\|\cdot\|$ is any norm.

Theorem (Upper bound of the border rank)

The border rank of a tensor $A(\ell, k)$ representing an ℓ -out-of-k function is at most min $\{\ell + 1, k - \ell + 1\}$.

Given a symmetric tensor representing an ℓ -out-of-k function our goal is to find another symmetric tensor:

• of the same order and the same dimensions

Given a symmetric tensor representing an ℓ -out-of-k function our goal is to find another symmetric tensor:

- of the same order and the same dimensions
- having symmetric rank at most $r = min\{\ell + 1, k \ell + 1\}$

Given a symmetric tensor representing an ℓ -out-of-k function our goal is to find another symmetric tensor:

- of the same order and the same dimensions
- having symmetric rank at most $r = min\{\ell + 1, k \ell + 1\}$
- that is a good approximation of the original tensor.

Given a symmetric tensor representing an ℓ -out-of-k function our goal is to find another symmetric tensor:

- of the same order and the same dimensions
- having symmetric rank at most $r = min\{\ell + 1, k \ell + 1\}$
- that is a good approximation of the original tensor.

We used a kind of stochastic hill-clibing algorithm.

$$\left(\begin{array}{cc} \left(\begin{array}{cc} 0\\1\end{array}\right) & \left(\begin{array}{cc} 1\\0\end{array}\right)\\ \left(\begin{array}{cc} 1\\0\end{array}\right) & \left(\begin{array}{cc} 0\\0\end{array}\right) \end{array}\right)$$

$$\begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{pmatrix}$$

$$\sim \frac{-0.19}{\exp(-15.75)} (1, \exp(-15.75)) \otimes \ldots \otimes (1, \exp(-15.75))$$

$$\begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{pmatrix} \\ \sim \frac{-0.19}{\exp(-15.75)} (1, \exp(-15.75)) \otimes \dots \otimes (1, \exp(-15.75)) \\ + \frac{1.19}{\exp(-13.90)} (1, \exp(-13.90)) \otimes \dots \otimes (1, \exp(-13.90))$$

$$\begin{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{pmatrix}$$

$$\sim \frac{-0.19}{\exp(-15.75)} (1, \exp(-15.75)) \otimes \dots \otimes (1, \exp(-15.75))$$

$$+ \frac{1.19}{\exp(-13.90)} (1, \exp(-13.90)) \otimes \dots \otimes (1, \exp(-13.90))$$

$$= \begin{pmatrix} \begin{pmatrix} 2.33 \cdot 10^{-10} \\ 1.0 \end{pmatrix} & \begin{pmatrix} 1.0 \\ 1.07 \cdot 10^{-6} \\ 9.96 \cdot 10^{-13} \end{pmatrix}$$

In the real world there is usually a noise that modifies functional relations between variables.

In the real world there is usually a noise that modifies functional relations between variables.

Tensor $N(\ell,k,p,q)$ represents an ℓ -out-of-k function with noisy inputs if it holds for $(i_1,\ldots,i_k)\in\{0,1\}^k$ that

In the real world there is usually a noise that modifies functional relations between variables.

Tensor $N(\ell, k, p, q)$ represents an ℓ -out-of-k function with noisy inputs if it holds for $(i_1, \ldots, i_k) \in \{0, 1\}^k$ that

$$\begin{split} N(\ell,k,p,q)_{i_1,i_2,\dots,i_k} \\ &= \sum_{(j_1,j_2,\dots,j_k) \in \{0,1\}^k} A_{j_1,j_2,\dots,j_k}(\ell,k) \cdot \prod_{n=1}^k M_{i_n,j_n}(p,q) \ , \end{split}$$

In the real world there is usually a noise that modifies functional relations between variables.

Tensor $N(\ell, k, p, q)$ represents an ℓ -out-of-k function with noisy inputs if it holds for $(i_1, \ldots, i_k) \in \{0, 1\}^k$ that

$$\begin{split} N(\ell,k,p,q)_{i_1,i_2,\dots,i_k} \\ &= \sum_{(j_1,j_2,\dots,j_k) \in \{0,1\}^k} A_{j_1,j_2,\dots,j_k}(\ell,k) \cdot \prod_{n=1}^k M_{i_n,j_n}(p,q) \ , \end{split}$$

where $A_{j_1,j_2,\dots,j_k}(\ell,k)$ represents the (exact) ℓ -out-of-k function,

In the real world there is usually a noise that modifies functional relations between variables.

Tensor $N(\ell, k, p, q)$ represents an ℓ -out-of-k function with noisy inputs if it holds for $(i_1, \dots, i_k) \in \{0, 1\}^k$ that

$$\begin{split} N(\ell,k,p,q)_{i_1,i_2,\dots,i_k} \\ &= \sum_{(j_1,j_2,\dots,j_k) \in \{0,1\}^k} A_{j_1,j_2,\dots,j_k}(\ell,k) \cdot \prod_{n=1}^k M_{i_n,j_n}(p,q) \ , \end{split}$$

where $A_{j_1,j_2,...,j_k}(\ell,k)$ represents the (exact) ℓ -out-of-k function, $M_{i_1,i_2}(p,q)$ are elements of matrix M(p,q) defined by

$$M_{i_n,j_n}(p,q) \ = \ \begin{cases} \ q & \text{if } j_n = 0 \text{ and } i_n = 0 \\ 1-q & \text{if } j_n = 1 \text{ and } i_n = 0 \\ 1-p & \text{if } j_n = 0 \text{ and } i_n = 1 \\ p & \text{if } j_n = 1 \text{ and } i_n = 1 \end{cases}$$

 $0 , <math>0 < q \le 1$ are parameters of the input noise.

We performed experiments with the game of Minesweeper for the 20×20 grid size.

We performed experiments with the game of Minesweeper for the 20×20 grid size.

We used a random selection of fields to be played and we assumed we never hit any of fifty mines during the game.

We performed experiments with the game of Minesweeper for the 20×20 grid size.

We used a random selection of fields to be played and we assumed we never hit any of fifty mines during the game.

At each of 350 steps of the game we created a Bayesian network

 the standard method consisting of moralization and triangulation steps and

We performed experiments with the game of Minesweeper for the 20×20 grid size.

We used a random selection of fields to be played and we assumed we never hit any of fifty mines during the game.

At each of 350 steps of the game we created a Bayesian network

- the standard method consisting of moralization and triangulation steps and
- the tensor rank-one decomposition applied to CPTs with number of parents higher than three (for CPTs with less than four parents we used the moralization) followed by the triangulation step.

We performed experiments with the game of Minesweeper for the 20×20 grid size.

We used a random selection of fields to be played and we assumed we never hit any of fifty mines during the game.

At each of 350 steps of the game we created a Bayesian network

- the standard method consisting of moralization and triangulation steps and
- the tensor rank-one decomposition applied to CPTs with number of parents higher than three (for CPTs with less than four parents we used the moralization) followed by the triangulation step.

In both networks we then used the lazy propagation method of Madsen and Jensen with the computations performed with lists of tables over the junction trees.

Results of experiments

Numerical experiments reveal that we can get a gain in the order of two magnitudes but at the expense of a certain loss of precision. See Figure.

