# Re-exam Probabilistic Reasoning

15 January 2020, 17:00 – 20:00

The exam consists of three problems, with independent subproblems; in total a maximum of 100 points is awarded. Read the questions very carefully and clearly explain your answers (in English and/or Dutch).

Good luck!

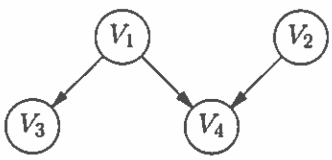
## Problem 1 (a: 5 pts., b: 10 pts., c: 10 pts., d: 10 pts., total: 35 pts.)

a. Consider using *Pearl's data fusion algorithm* and its computation rules (see attached formula sheet) on a given Bayesian network $\mathcal{B}$. Is the following statement true or false?

"For all non-observed nodes $V_i$ in $\mathcal{B}$, it holds that $\lambda(v_i) + \lambda(\neg v_i) = 1$."

Clearly explain your answer.

b. Consider the Bayesian network $\mathcal{B} = (G, \Gamma)$ with the following acyclic digraph $G$ and assessment functions $\Gamma$ (complements omitted):




$\gamma(v_1) = 0.5$

$\gamma(v_2) = 0.8$

$\gamma(v_3 \mid v_1) = 0.2$
$\gamma(v_3 \mid \neg v_1) = 0.6$

$\gamma(v_4 \mid v_1 \wedge v_2) = 0.9$
$\gamma(v_4 \mid \neg v_1 \wedge v_2) = 0.6$
$\gamma(v_4 \mid v_1 \wedge \neg v_2) = 0.4$
$\gamma(v_4 \mid \neg v_1 \wedge \neg v_2) = 0.8$

Suppose that the evidence $V_4 = true$ is entered into $\mathcal{B}$. Illustrate Pearl's algorithm by computing the posterior probability $\Pr^{v_4}(v_2)$ from the network. Explicitly list the values of all separate and compound causal and diagnostic parameters used; if you use properties other than Pearl's computation rules, explicitly indicate and explain these.

c. Consider again the Bayesian network $\mathcal{B}$ from part b. Suppose that, after propagating the evidence $V_4 = true$, the additional evidence $V_3 = false$ is entered. Clearly list all separate and compound causal and diagnostic parameters which will change value as a consequence of propagating this additional evidence. For each of these parameters, describe which of their terms will *change value or will be newly inserted*; it is not necessary to do the actual computations involved.

d. Consider again the network $\mathcal{B}$ from part b. Can the probability $p = \Pr^{v_4}(v_1 \vee v_2)$ be computed from $\mathcal{B}$ by means of Pearl's data fusion algorithm ? Choose one of the following possible answers, and clearly explain why your choice is the correct answer:

**I** Yes, $p$ can be computed directly from $\mathcal{B}$ by a single application of Pearl's algorithm.

**II** Yes, $p$ can be computed indirectly from $\mathcal{B}$ by combining the results from multiple consecutive applications of Pearl's algorithm.

**III** No, Pearl's algorithm cannot be used since it cannot yield any probabilities from which $p$ can be established.

## Problem 2 (a: 10 pts., b: 10 pts., c: 10 pts., total: 30 pts.)

Suppose you want to construct a Bayesian network from a data set $D$ using a learning algorithm that combines a search heuristic with the MDL quality measure (see formula sheet). Here we consider two different search heuristics:

- the $B$ search heuristic: starts with an *empty* graph and subsequently *adds* arcs that result in the largest increase in quality;
- the $\overline{B}$ search heuristic: starts with a *complete* acyclic directed graph and *removes* arcs that result in the largest increase in quality.

Consider the following data set $D$ over binary-valued random variables $V = \{V_1, V_2, V_3\}$:

| | | |
|---|---|---|
| $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge \neg v_2 \wedge v_3$ |
| $v_1 \wedge \neg v_2 \wedge v_3$ | $\neg v_1 \wedge v_2 \wedge v_3$ | $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ |
| $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge v_2 \wedge v_3$ | $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ |
| $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge v_2 \wedge \neg v_3$ | $\neg v_1 \wedge v_2 \wedge \neg v_3$ |
| $v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge \neg v_2 \wedge \neg v_3$ | $\neg v_1 \wedge v_2 \wedge \neg v_3$ |

There is no further information available about $V_1, V_2$ and $V_3$, and how they are related.

a. Suppose that for constructing the digraph of the network, the $\overline{B}$ search heuristic is used. Let the heuristic start with the following complete graph:

$$G = (\{V_1, V_2, V_3\}, \{V_1 \to V_2, V_1 \to V_3, V_2 \to V_3\})$$

The node quality for node $V_3$ in this graph equals: $-5.7247$ (using base-10 log).

Compute the change in quality due to removing arc $V_1 \to V_3$ from $G$; assume that $P(G)$ is constant. Will the arc indeed be removed? Clearly explain your answers.

b. Now suppose we have used both search heuristics and compare their results. Let $G$ denote the acyclic digraph that results using the $B$ search heuristic, and let $\overline{G}$ denote the digraph resulting from the $\overline{B}$ heuristic.

Will, in general, $G = \overline{G}$? Clearly explain your answer.

c. Suppose that in the learned network $\rho(V_2) = \emptyset$. A domain expert, however, indicates that she expects an arc $V_1 \to V_2$. You wonder whether you could use an $n$-way sensitivity analysis to simulate the possible differences between presence or absence of this arc on the output of the network.

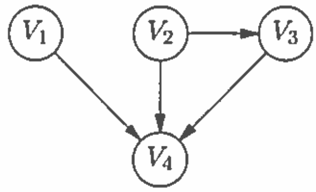
Is it indeed possible to employ a sensitivity analysis for this purpose? If so, clearly describe how you would do that. If not, clearly explain the reason(s).

2018-2019
4 Network science
3 Big data/ec
2 ~~ec~~ pr
1 data mining

2017-2018
1 ADS
2 CPD/A&N
3 ...
4

**Problem 3** (a: 10 pts., b: 10 pts., c: 10 pts., d: 5 pts., total: 35 pts.)

a. Consider the following directed acyclic digraph $G$:




and the independence relation $I$ defined on $V = \{V_1, V_2, V_3, V_4\}$ by the following statements (and any statements that can be derived from them using the axioms):

$$I(\{V_1\}, \emptyset, \{V_2, V_3\}) \quad \text{and} \quad I(\{V_2\}, \{V_4\}, \{V_3\})$$

Graph $G$ is an I-map of independence relation $I$. Suppose we construct graph $G^-$ by removing arc $V_2 \rightarrow V_3$ from $G$. Which of the following statements is true for $G^-$? Explain your answer.

**I** $G^-$ is an I-map for $I$, but *not* a D-map
**II** $G^-$ is *not* an I-map for $I$, but it is a D-map
**III** $G^-$ is neither an I-map nor a D-map
**IV** $G^-$ is a P-map

b. Consider Bayesian network $\mathcal{B} = (G^+, \Gamma)$ where graph $G^+$ is the result of adding arc $V_3 \rightarrow V_1$ to graph $G$ from part a.

- Give a minimal loop cutset for graph $G^+$.
- Comment on the convenience of your choice of loop cutset for computing the probability $\Pr^{v_1,v_2}(v_4)$ from $\mathcal{B}$.

Clearly explain your considerations.

c. Consider Bayesian network $\mathcal{B} = (G^-, \Gamma)$ where $G^-$ is the graph from part a (i.e without the arc $V_2 \rightarrow V_3$). $\Gamma$ is (partially) specified by the following (complements omitted):

$$\gamma(v_1) = 0.4 \qquad \gamma(v_4 \mid \neg v_1 \wedge \neg v_2 \wedge \neg v_3) = 0.0$$
$$\gamma(v_4 \mid v_1 \wedge \neg v_2 \wedge \neg v_3) = 0.4$$
$$\gamma(v_2) = 0.5 \qquad \gamma(v_4 \mid \neg v_1 \wedge v_2 \wedge \neg v_3) = 0.7$$
$$\gamma(v_4 \mid \neg v_1 \wedge \neg v_2 \wedge v_3) = 0.8$$
$$\gamma(v_3) = 0.6$$

Suppose the interaction between node $V_4$ and its parents $V_1, V_2$ and $V_3$ is modelled by a 'noisy-or gate'. Complete the assessment function for node $V_4$. Explain your answers.

d. The literature defines a measure of *data conflict* for observations $e_1, \ldots, e_m$ for a set of $m$ variables, based upon the idea that observations should originate from a coherent case and therefore correlate positively. More specifically, this measure is defined as:

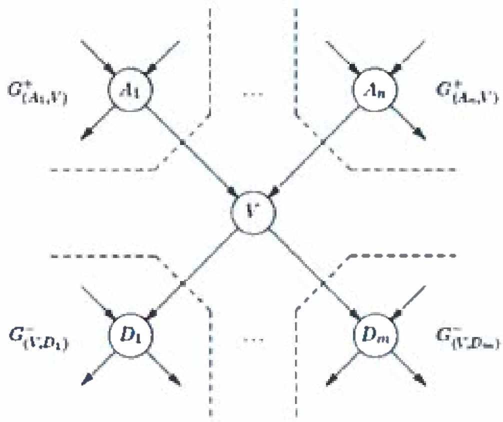
$$\mathsf{confl}(e_1, \ldots, e_m) = \log_2 \frac{\Pr(e_1) \cdot \ldots \cdot \Pr(e_m)}{\Pr(e_1 \wedge \ldots \wedge e_m)}$$

- Determine $\mathsf{confl}(v_1, \neg v_2, \neg v_3)$ for network $\mathcal{B} = (G^-, \Gamma)$ from part c.
- Comment on the suitability of this measure in the context of a Bayesian network in which the observable variables are modelled as causes in a noisy-or model.

# Formulas (Probabilistic Reasoning Exam Jan 2020)

## Pearl in a singly connected digraph

Consider a Bayesian network $B = (G, \Gamma)$ and a node $V$ in $G$ with direct ancestors (parents) $\rho(V) = \{A_1, \ldots, A_n\}$ and direct descendants (children) $\sigma(V) = \{D_1, \ldots D_m\}$.




To compute its (prior or posterior) probability distribution with Pearl's algorithm, node $V$ uses *data fusion*: $\alpha \cdot \pi(V) \cdot \lambda(V)$ and computes the following parameters for all $c_V$ and $c_{A_i}$:

$$\pi(V) = \sum_{c_{\rho(V)}} \Big(\gamma(V \mid c_{\rho(V)}) \cdot \prod_{i=1,\ldots,n} \pi_V^{A_i}(c_{A_i})\Big)$$

$$\lambda(V) = \prod_{j=1,\ldots,m} \lambda_{D_j}^V(V)$$

$$\pi_{D_j}^V(V) = \alpha \cdot \pi(V) \cdot \prod_{\substack{k=1,\ldots,m \\ k \neq j}} \lambda_{D_k}^V(V)$$

$$\lambda_V^{A_i}(A_i) = \alpha \cdot \sum_{c_V} \lambda(c_V) \cdot \sum_{c_{\rho(V)\setminus\{A_i\}}} \Big(\gamma(c_V \mid c_{\rho(V)\setminus\{A_i\}} \wedge A_i) \cdot \prod_{\substack{k=1,\ldots,n \\ k \neq i}} \pi_V^{A_k}(c_{A_k})\Big)$$

## The MDL quality measure

Let $G = (\mathbf{V}_G, \mathbf{A}_G)$ be an acyclic digraph and let $\mathbf{D}$ be a dataset over $N$ cases. Let $P(G)$ be a probability distribution over the set of acyclic graphs with node set $\mathbf{V}$. Then, the MDL quality measure for graph $G$ is given by

$$Q_{MDL}(G, \mathbf{D}) = \log P(G) - N \cdot H(G, \mathbf{D}) - \frac{1}{2}\log(N) \cdot \sum_{V_i \in V} 2^{|\rho(V_i)|}$$

$$= \log P(G) + \sum_{V_i \in V} q(V_i, \rho(V_i), \mathbf{D})$$

where $q(V_i, \rho(V_i), \mathbf{D})$ is the quality of node $V_i$ and

$$-N \cdot H(G, \mathbf{D}) = \sum_{V_i \in V} \sum_{c_{V_i}} \sum_{c_{\rho(V_i)}} N(c_{V_i} \wedge c_{\rho(V_i)}) \cdot \log\Big(\frac{N(c_{V_i} \wedge c_{\rho(V_i)})}{N(c_{\rho(V_i)})}\Big) \qquad (0 \cdot \log\frac{0}{x} = 0, \text{ even if } x = 0)$$