Exam Probabilistic Reasoning

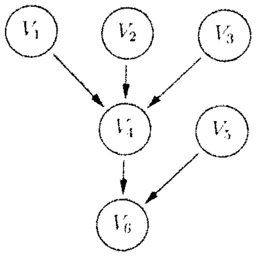
6 November 2013

The exam consists of three problems, for each of which the number of credits per question is given. In total, a maximum of 100 credits is awarded. Read the questions very carefully; you may answer them in English and/or Dutch. Be sure to clearly explain your answers!

Good luck!

Problem 1 (a: 10 pts., b: 10 pts., c: 10 pts., total: 30 pts.)

Consider a probabilistic network $B = (G, \Gamma)$, where $G = (V(G), A(G))$ is the following acyclic digraph and $\Gamma = \{\gamma_{V_i} \mid V_i \in V(G)\}$ is given by:



$\gamma(v_1) = 0.9$ $\qquad$ $\gamma(v_2) = 0.4$

$\gamma(v_3) = 0.5$ $\qquad$ $\gamma(v_5) = 0.7$

$\gamma(v_4 \mid v_1 \wedge \neg v_2 \wedge \neg v_3) = 0$
$\gamma(v_4 \mid \neg v_1 \wedge v_2 \wedge \neg v_3) = 0.5$
$\gamma(v_4 \mid \neg v_1 \wedge \neg v_2 \wedge v_3) = 0.2$

$\gamma(v_6 \mid v_4 \wedge v_5) = 0.1$
$\gamma(v_6 \mid \neg v_4 \wedge v_5) = 0.3$
$\gamma(v_6 \mid v_4 \wedge \neg v_5) = 0.7$
$\gamma(v_6 \mid \neg v_4 \wedge \neg v_5) = 0.6$

Variables V_1, V_2 and V_3 have a disjunctive interaction effect on variable V_4. To capture this effect, the assessment function for node V_4 is based on the 'noisy-or gate'.

a. Complete the assessment function $\gamma(V_4)$ for node V_4. Explain your answers.

Let Pr be the probability distribution defined by probabilistic network B. Now, consider the five computation rules of *Pearl*'s data fusion algorithm, given in the attached formula sheet.

b. Illustrate Pearl's algorithm by computing the prior probabilities $\Pr(v_4)$ and $\Pr(\neg v_4)$ from network B.

   Clearly indicate which probabilities are computed, which messages are passed and how they are computed; explicitly mention all assumptions you make.

For the next question you may assume that you have an inference algorithm at your disposal (e.g. Pearl) which can compute probabilities of the form $\Pr(c_{V_i} \mid c_E)$ $(V_i \in V(G),\ E \subseteq V(G))$.

c. You are interested in computing the probability $\Pr(v_1 \wedge \neg v_4 \wedge v_6)$. Clearly explain *how* you can compute this specific probability *efficiently* from network B.

   Note: you don't have to perform the actual computation, just explain the approach you would take.

Problem 2 (a: 10 pts., b: 10 pts., c: 15 pts., total: 35 pts.)

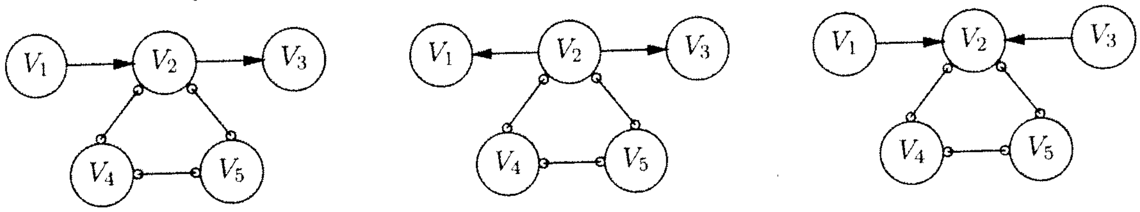
Consider a probabilistic network $B = (G, \Gamma)$, a node of interest $T \in V(G)$ and a set of observed nodes $E \subseteq V(G)$ such that $T \notin E$. In a probabilistic network we can identify nodes that are *barren* with respect to T and E: a node $X \in V(G)$ is called *barren* if

$X \neq T$ and $X \notin E$, and all its descendants $\sigma^*(X)$ are barren

a. Consider the probabilistic network from **Problem 1**. Let $T = V_2$ and let $E = \varnothing$. Give all nodes that are barren with respect to T and E. Explain your answer.

b. To speed up inference, a probabilistic network can be *pruned* to a computationally equivalent network for computing a specific distribution $\Pr(T \mid c_E)$. To this end, all d-separated nodes and all barren nodes with respect to T and E are removed. Clearly explain why these two sets of nodes can be safely removed for computing $\Pr(T \mid c_E)$.

Consider the following graphical structures, where each o indicates the presence or absence of an arrowhead: by replacing every connection o— o with an arc, a directed graph results; we consider only *acyclic* versions.



The graphs all include two connections between V_1 and V_3: the *simple* chain V_1 o—o V_2 o—o V_3 and the 'loopy' chain V_1o—o V_2 o—o V_4 o—o V_5 o—o V_2 o—o V_3, which passes node V_2 twice. Recall that in the context of *blocking* and *d-separation* it is sufficient to consider *simple* chains only, where every node is included at most once.

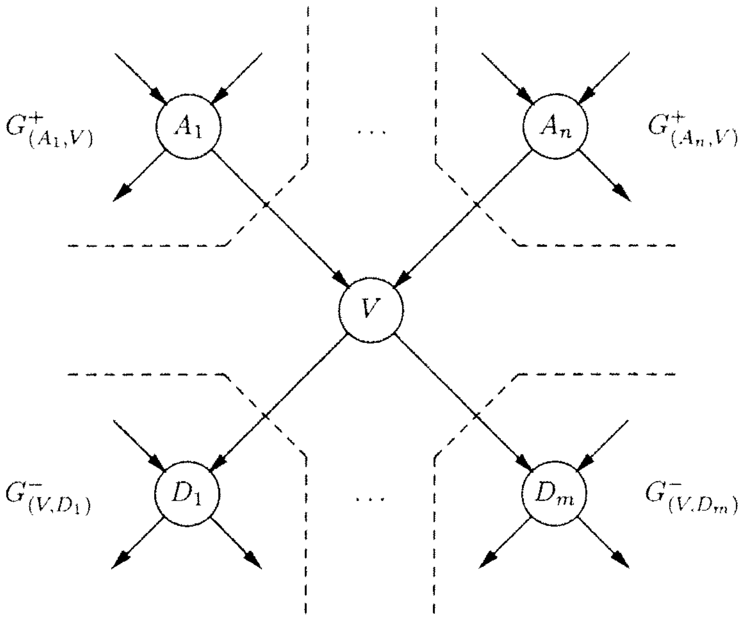
c. Clearly demonstrate for all three graphs described above that the following holds regardless of the direction of the arcs for V_2 o—o V_4, V_4 o—o V_5 and V_5 o—o V_2:

   if the simple chain between nodes V_1 and V_3 is blocked then the 'loopy' chain between these nodes is also blocked.

   Hint: You may disregard all situations with evidence for V_1 or V_3.

Formulas (Probabilistic Reasoning Exam Nov 2013)

Pearl in a singly connected digraph



Consider a node V in a probabilistic network $B = (G, \Gamma)$. Let $\rho(V) = \{A_1, \ldots, A_n\}$ be the set of direct ancestors (parents) of V in G, and let $\sigma(V) = \{D_1, \ldots D_m\}$ be the set of its direct descendants (children). With Pearl's algorithm, node V computes the following parameters:

$$
\begin{aligned}
\pi(V) &= \sum_{c_{\rho(V)}} \Big(\gamma\left(V \mid c_{\rho(V)}\right) \cdot \prod_{i=1,\ldots,n} \pi_V^{A_i}(c_{A_i}) \Big) \\
\lambda(V) &= \prod_{j=1,\ldots,m} \lambda_{D_j}^{V}(V) \\
\pi_{D_j}^{V}(V) &= \alpha \cdot \pi(V) \cdot \prod_{\substack{k=1,\ldots,m \\ k \neq j}} \lambda_{D_k}^{V}(V) \\
\lambda_V^{A_i}(A_i) &= \alpha \cdot \sum_{c_V} \lambda(c_V) \cdot \sum_{c_{\rho(V)\setminus\{A_i\}}} \Big(\gamma\left(c_V \mid c_{\rho(V)\setminus\{A_i\}} \wedge A_i\right) \cdot \prod_{\substack{k=1,\ldots,n \\ k \neq i}} \pi_V^{A_k}(c_{A_k}) \Big)
\end{aligned}
$$

in order to perform *data fusion*: $\alpha \cdot \pi(V) \cdot \lambda(V)$, which results in V's (prior or posterior) probability distribution.

Problem 3 (a: 10 pts., b: 10 pts., c: 15 pts., total: 35 pts.)

Suppose that you are constructing a probabilistic network for a certain domain of application. For quantifying the network, a number of probabilities have to be assessed by a human expert. After full quantification, you will perform a sensitivity analysis to investigate the robustness of your network.

a. Describe two different methods for eliciting probabilities from human experts. For each of the two methods, provide a benefit and a drawback.

Suppose the graph of your probabilistic network is the same as that from **Problem 1**.

b. Give the *Sensitivity set* $S^{\{V_4,V_6\}}(V_3)$ for node of interest V_3 given evidence for nodes V_4 and V_6. Clearly explain your answer.

Consider a sensitivity function $f(x)$ for some output probability and some parameter $x = \gamma_V(v_i)$ of variable V with $n > 2$ values. Upon varying x, the values of the assessment function γ_V for the other $n-1$ values of variable V have to be co-varied. To this end, a *proportional* co-variation scheme is used in which the proportion of the remaining mass $1-x$ that is assigned to $\gamma_V(v_j)$, $j \neq i$, is kept constant.

Different co-variation schemes adhere to different properties. Two useful properties are the *order-preserving* property and the *impossibility-preserving* property. A co-variation scheme is called impossibility-preserving if any $\gamma_V(v_j) = 0$, $j \neq i$, remains zero upon co-variation. Suppose the values of V are ordered according to the values of γ_V as specified in the network, i.e. $\gamma_V(v_1) \leq \ldots \leq \gamma_V(v_n)$; a co-variation scheme is called order-preserving if this ordering is preserved during co-variation.

c. Indicate which of the two properties described above are properties of the proportional co-variation scheme:

   I. both order-preserving and impossibility-preserving
   II. order-preserving, but not impossibility-preserving
   III. not order-preserving, yet impossibility-preserving
   IV. neither order-preserving, nor impossibility-preserving

   For each of the properties provide a clear proof or counter-example.