Solutions

Examination in Bayesian Networks **SESSION 2009–2010**

12th May, 2010

Question 1

a. We start with $P(c)$:

$$\begin{aligned} P(c) &= \sum_{A,B} P(c \mid A, B)P(A)P(B) \\ &= 0.8 \cdot 1 \cdot 0.5 + 0 + 0.3 \cdot 1 \cdot 0.5 + 0 \\ &= 0.55 \end{aligned}$$

Next we compute $P(a \mid c)$:

$$\begin{aligned} P(a \mid c) &= P(c \mid a)P(a)/P(c) \quad \text{by Bayes' rule} \\ &= \sum_{B} P(c \mid a, B)P(B \mid a)P(a)/P(c) \\ &= \sum_{B} P(c \mid a, B)P(B)P(a)/P(c) \quad \text{since } A \perp\!\!\!\perp_G B \mid \varnothing \\ &= (0.8 \cdot 0.5 \cdot 1 + 0.3 \cdot 0.5 \cdot 1)/0.55 \\ &= 1 \end{aligned}$$

We have that $P(a \mid d) = P(a) = 1$, as $A \perp\!\!\!\perp_G D \mid \varnothing$. Finally,

$$\begin{aligned} P(a, \neg b \mid c) &= P(c \mid a, \neg b)P(a)P(\neg b)/P(c) \\ &= 0.3 \cdot 1 \cdot 0.5/0.55 \\ &= 3/11 \end{aligned}$$

b. The number of probabilistic parameters you need to specify the family of probability distributions

$$P(Y \mid X_1, \dots, X_n)$$

is an exponential function of n. Causal independence can bring this back to a linear number of parameters of n. Another method that may help is devorcing multiple parents.

c. Using marginalisation and conditioning, it follows that

$$P(J) = \sum_{I_C, I_D} P(J \mid I_C, I_D)P(I_C)P(I_D) \tag{1}$$

as I_C and I_D are independent. The probability distribution $P(J \mid I_C, I_D)$ is assumed to model a logical OR, i.e.

$$P(j \mid I_C, I_D) = \begin{cases} 0 & \text{if } I_C = I_D = \perp \\ 1 & \text{otherwise} \end{cases}$$

Now, we have

$$\begin{aligned}
P(i_D) &= \sum_D P(i_D \mid D)P(D) \\
&= P(i_D \mid d)P(d) \quad \text{as } P(i_D \mid \neg d) = 0 \\
&= 0.3 \cdot 0.3 = 0.09
\end{aligned}$$

For variable C we already have from 1(a) that $P(c) = 0.55$. Finally,

$$\begin{aligned}
P(i_C) &= \sum_C P(i_C \mid C)P(C) \\
&= P(i_C \mid C)P(C) \quad \text{again } P(i_C \mid \neg c) = 0 \\
&= 0.5 \cdot 0.55 = 0.275
\end{aligned}$$

Substituting these results into equation (1) yields:

$$\begin{aligned}
P(j) &= 1 - P(\neg i_C)P(\neg i_D) \\
&= 1 - 0.725 \cdot 0.91 \\
&\approx 0.34
\end{aligned}$$

d. Meaning of XOR: two causes exclude each other, so only the combinations $i_C, \neg i_D$ and $\neg i_C, i_D$ come into the equation for $P(j)$. Computation of $P(j)$:

$$\begin{aligned}
P(j) &= P(i_C)P(\neg i_D) + P(\neg i_C)P(i_D) \\
&= 0.275 \cdot 0.91 + 0.725 \cdot 0.09 \\
&\approx 0.32
\end{aligned}$$

Question 2

a. It is a nonminimal I-map as the arc between V_1 and V_3 is redundant because $P(V_3 \mid V_1, V_2) = P(V_3 \mid V_2)$. The undirected I-map of P is shown in Figure 1.

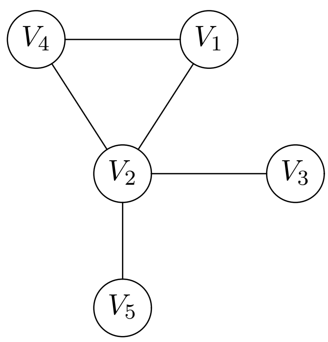

Figure 1: Undirected I-map.

b. Some possibilities:

(1) $V_4 \perp\!\!\!\perp_G V_1 \mid \varnothing$

(2) $V_4 \perp\!\!\!\perp_G V_5 \mid V_2$

(3) $V_1 \perp\!\!\!\perp_G V_3 \mid V_2$

(4) $V_1 \perp\!\!\!\perp_G V_5 \mid V_2$

(5) $V_3 \perp\!\!\!\perp_G V_5 \mid \{V_2, V_1\}$

c. The assumptions of independence of cases in the database D are rarely fully justified. Similarly, it may not always the case that the cases come from the same process and are thus from the same probability distribution. Likelihood would yield the same value for Markov equivalent Bayesian networks, as these represent exactly the same probability distribution, and this definition is taken to compute the likelihood.

d. The disadvantage of greedy search is that it cannot later reconsider the choices made for the next likely graph structure that is being explored in the search process. An alternative is to allow returning to states that were not yet considered by keeping track of a list of maximum size of not yet considered structures. The most unrestricted version of such an algorithm is the A^* algorithm, that allows to explore the whole search space exhaustively.