ARTIFICIAL INTELLIGENCE (E016350B)
GHENT UNIVERSITY
AY 2024/2025

Prof. Aleksandra Pizurica
Assistant: Nicolas Vercheval
E-mail: ai@lists.ugent.be

Solutions: Inference in Bayesian networks

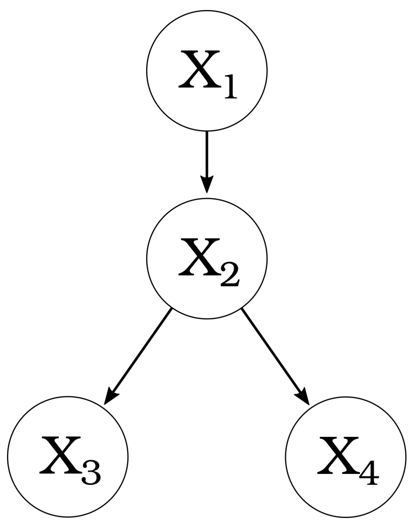
1. Given the Bayesian network below, calculate marginal and conditional probabilities $P(\neg x_3)$ and $P(x_2 \mid \neg x_3)$ by using:

   (a) the method of **inference by enumeration**,

   (b) the method of **variable elimination**.

$\mathbf{P}(X_1)$	
x_1	0.4
$\neg x_1$	0.6

$\mathbf{P}(X_3 \mid x_2)$		
x_2	x_3	0.2
x_2	$\neg x_3$	0.8
$\neg x_2$	x_3	0.3
$\neg x_2$	$\neg x_3$	0.7



$\mathbf{P}(X_2 \mid X_1)$		
x_1	x_2	0.8
x_1	$\neg x_2$	0.2
$\neg x_1$	x_2	0.5
$\neg x_1$	$\neg x_2$	0.5

$\mathbf{P}(X_4 \mid X_2)$		
x_2	x_4	0.8
x_2	$\neg x_4$	0.2
$\neg x_2$	x_4	0.5
$\neg x_2$	$\neg x_4$	0.5

Solution:

(a) Inference by enumeration sums the joint probabilities of atomic events. For a typical query, we express

$$\mathbf{P}(Q \mid \mathbf{e}) = \alpha \mathbf{P}(Q, \mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(Q, \mathbf{e}, \mathbf{h})$$

and we express the joint probability $\mathbf{P}(Q, \mathbf{e}, \mathbf{h})$ using the particular structure of the probabilistic model. For a Bayesian network consisting of random variables $X_1, \ldots, X_n$, this joint probability is:

$$\mathbf{P}(X_1, \ldots, X_n) = \mathbf{P}(X_1 \mid parents(X_1)) \cdot \ldots \cdot \mathbf{P}(X_n \mid parents(X_n)).$$

Some of these random variables will be the query variables, others evidence or hidden variables. For the given Bayesian network, we have the following:

$$
\begin{aligned}
P(\neg x_3) &= \sum_{x_1,x_2,x_4} P(x_1, x_2, \neg x_3, x_4) \\
&= \sum_{x_1,x_2,x_4} P(x_1)P(x_2|x_1)P(\neg x_3 \mid x_2)P(x_4|x_2) \\
&= P(x_1)P(x_2 \mid x_1)P(\neg x_3 \mid x_2)P(x_4 \mid x_2) \\
&+ P(x_1)P(x_2 \mid x_1)P(\neg x_3 \mid x_2)P(\neg x_4 \mid x_2) \\
&+ P(x_1)P(\neg x_2 \mid x_1)P(\neg x_3 \mid \neg x_2)P(x_4 \mid \neg x_2) \\
&+ P(x_1)P(\neg x_2 \mid x_1)P(\neg x_3 \mid \neg x_2)P(\neg x_4 \mid \neg x_2) \\
&+ P(\neg x_1)P(x_2 \mid \neg x_1)P(\neg x_3 \mid x_2)P(x_4 \mid x_2) \\
&+ P(\neg x_1)P(x_2 \mid \neg x_1)P(\neg x_3 \mid x_2)P(\neg x_4 \mid x_2) \\
&+ P(\neg x_1)P(\neg x_2 \mid \neg x_1)P(\neg x_3 \mid \neg x_2)P(x_4 \mid \neg x_2) \\
&+ P(\neg x_1)P(\neg x_2 \mid \neg x_1)P(\neg x_3 \mid \neg x_2)P(\neg x_4 \mid \neg x_2) \\
&= 0.4 \cdot 0.8 \cdot 0.8 \cdot 0.8 + 0.4 \cdot 0.8 \cdot 0.8 \cdot 0.2 + 0.4 \cdot 0.2 \cdot 0.7 \cdot 0.5 \\
&+ 0.4 \cdot 0.2 \cdot 0.7 \cdot 0.5 + 0.6 \cdot 0.5 \cdot 0.8 \cdot 0.8 + 0.6 \cdot 0.5 \cdot 0.8 \cdot 0.2 \\
&+ 0.6 \cdot 0.5 \cdot 0.7 \cdot 0.5 + 0.6 \cdot 0.5 \cdot 0.7 \cdot 0.5 \\
&= 0.2048 + 0.0512 + 0.028 + 0.028 + 0.192 + 0.048 + 0.105 + 0.105 \\
&= 0.762.
\end{aligned}
$$

Similarly, we obtain:

$$
\begin{aligned}
P(x_2, \neg x_3) &= \sum_{x_1,x_4} P(x_1, x_2, \neg x_3, x_4) \\
&= \sum_{x_1,x_4} P(x_1)P(x_2 \mid x_1)P(\neg x_3 \mid x_2)P(x_4 \mid x_2) \\
&= P(x_1)P(x_2 \mid x_1)P(\neg x_3 \mid x_2)P(x_4 \mid x_2) \\
&+ P(x_1)P(x_2 \mid x_1)P(\neg x_3 \mid x_2)P(\neg x_4 \mid x_2) \\
&+ P(\neg x_1)P(x_2 \neg\neg x_1)P(\neg x_3 \mid x_2)P(x_4 \mid x_2) \\
&+ P(\neg x_1)P(x_2 \neg x_1)P(\neg x_3 \mid x_2)P(\neg x_4 \mid x_2) \\
&= 0.4 \cdot 0.8 \cdot 0.8 \cdot 0.8 + 0.4 \cdot 0.8 \cdot 0.8 \cdot 0.2 \\
&+ 0.6 \cdot 0.5 \cdot 0.8 \cdot 0.8 + 0.6 \cdot 0.5 \cdot 0.8 \cdot 0.2 \\
&= 0.2048 + 0.0512 + 0.192 + 0.048 = 0.496
\end{aligned}
$$

and

$$
P(x_2 \mid \neg x_3) = \frac{P(x_2, \neg x_3)}{P(\neg x_3)} = \frac{0.496}{0.762} = 0.6509.
$$

The method repeats many calculations. It is a routine and easily formalized algorithm, but computationally expensive. Its complexity is exponential in the number of variables.

Remark: Note that we could employ the properties of directed graphical model to manually simplify inference by enumeration. When calculating $P(\neg x_3)$ and also $P(x_2 \mid \neg x_3)$, X_4 is a leaf that is neither a query nor evidence. This means that it can be eliminated without changing the target probabilities. *(Do this as an exercise!).*

(b) Variable elimination avoids repeated calculations. The idea is simply to do each calculation once and store it for later use. We again express a typical query as

$$\mathbf{P}(Q \mid \mathbf{e}) = \alpha \mathbf{P}(Q, \mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(Q, \mathbf{e}, \mathbf{h})$$

and we again express the joint probability $\mathbf{P}(Q, \mathbf{e}, \mathbf{h})$ using the structure (conditional independences) in the particular network. But now we will group the resulting terms into factors that will be calculated once and (as intermediate results) stored for further use. We have

$$\begin{aligned} P(\neg x_3) &= \sum_{x_1,x_2,x_4} P(x_1, x_2, \neg x_3, x_4) \\ &= \sum_{x_1,x_2,x_4} P(x_1)P(x_2|x_1)P(\neg x_3 \mid x_2)P(x_4|x_2) \\ &= \sum_{x_1} \underbrace{P(x_1)}_{\mathbf{f}_1(X_1)} \sum_{x_2} \underbrace{P(x_2|x_1)}_{\mathbf{f}_2(X_1,X_2)} \underbrace{P(\neg x_3 \mid x_2)}_{\mathbf{f}_3(X_2)} \underbrace{\sum_{x_4} P(x_4|x_2)}_{1} \end{aligned}$$

All the three factors are given by the corresponding CPT's in the problem description, e.g., $\mathbf{f}_1(X_1) = \langle 0.4, 0.6 \rangle$ and $\mathbf{f}_3(X_2) = \langle 0.8, 0.7 \rangle$ We can also choose this ordering:

$$P(\neg x_3) = \sum_{x_2} \underbrace{P(\neg x_3 \mid x_2)}_{\mathbf{f}_3(X_2)} \sum_{x_1} \underbrace{P(x_1)}_{\mathbf{f}_1(X_1)} \underbrace{P(x_2|x_1)}_{\mathbf{f}_2(X_1,X_2)} \underbrace{\sum_{x_4} P(x_4|x_2)}_{1}$$

Both must give the same result. With this second one, we form the factor

$$\mathbf{f}_4(X_2) = \sum_{x_1} \mathbf{f}_1(X_1) \times \mathbf{f}_2(X_1, X_2) = 0.4 \cdot \langle 0.8, 0.2 \rangle + 0.6 \cdot \langle 0.5, 0.5 \rangle = \langle 0.62, 0.38 \rangle$$

Then we need just one more step:

$$P(\neg x_3) = \sum_{x_2} \mathbf{f}_3(X_2) \times \mathbf{f}_4(X_2) = 0.8 \cdot 0.62 + 0.7 \cdot 0.38 = 0.762.$$

- $P(x_2 \mid \neg x_3) =?$

Let us express first

$$\begin{aligned}
\mathbf{P}(X_2|\neg x_3) &= \alpha \mathbf{P}(X_2, \neg x_3) = \alpha \sum_{x_1} \sum_{x_4} P(x_1)\mathbf{P}(X_2|x_1)\mathbf{P}(\neg x_3|X_2)\mathbf{P}(x_4|X_2) \\
&= \alpha \sum_{x_1} \underbrace{P(x_1)}_{\mathbf{f}_1(X_1)} \underbrace{\mathbf{P}(X_2|x_1)}_{\mathbf{f}_2(X_1,X_2)} \underbrace{\mathbf{P}(\neg x_3|X_2)}_{\mathbf{f}_3(X_2)} \underbrace{\sum_{x_4} \mathbf{P}(x_4|X_2)}_{\langle 1,1 \rangle} \\
&= \alpha \underbrace{\mathbf{P}(\neg x_3|X_2)}_{\mathbf{f}_3(X_2)} \underbrace{\sum_{x_1} \underbrace{P(x_1)}_{\mathbf{f}_1(X_1)} \underbrace{\mathbf{P}(X_2|x_1)}_{\mathbf{f}_2(X_1,X_2)}}_{\mathbf{f}_4(X_2)} = \alpha\ \mathbf{f}_4(X_2) \times \mathbf{f}_3(X_2) \\
&= \alpha \langle 0.62, 0.38 \rangle \langle 0.8, 0.7 \rangle = \alpha \langle 0.496, 0.266 \rangle = \langle 0.6509, 0.3491 \rangle
\end{aligned}$$

Note that all the factors $\mathbf{f}_1(X_1)$, $\mathbf{f}_2(X_1, X_2)$ and $\mathbf{f}_3(X_2)$ are here the same vectors as in the previous calculation (for the query $P(\neg x_3)$) although in that previous case we needed both values of X_2 because of the summation over x_2 and here because we kept X_2 as a variable. Hence, $\sum_{x_1} \mathbf{f}_1(X_1) \times \mathbf{f}_2(X_1, X_2)$ is the same factor $\mathbf{f}_4(X_2) = \langle 0.62, 0.38 \rangle$ that we already calculated before, so we just re-used it here.

Thus, we have that $P(x_2 \mid \neg x_3) = 0.6509$.

2. Suppose a burglary alarm which can function in its basic mode and with an additional ultrasonic sensor (which can be turned on or off). The network in Figure 2 represents a complete burglar alarm and notification system, with the following binary random variables: B (burglary happens when $B = 1$), U (ultrasonic sensor is on when $U = 1$), A (alarm sounds when $A = 1$) and N (neighbor calls when $N = 1$).

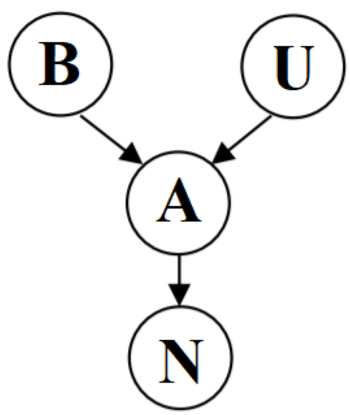


Figure 1: Bayesian network for a simplified burglar alarm system.

When the ultrasonic sensor is active, the probability that alarm is properly activated by burglary is increased from x to y, but the probability of false alarm is also increased, from ε to 2ε. (When the ultrasonic sensor is not active, the probability of false alarm is ε, and with ultrasonic sensor on, the false alarm appears with probability 2ε). Prior probability of burglary in the neighborhood where the house is located is b and the probability that the ultrasonic sensor is active is u.

(a) Write the joint probability for the network in Fig. 2 (a). Write also a table, which displays conditional probabilities of A given B and U.

(b) Express the probability that there is a burglary if alarm goes on in terms of x, y, u, ε and b.

(c) Suppose that the neighbor calls to report the alarm with probability n when the alarm is on and never when the alarm is off. Express the probability that there is a burglary if the neighbor calls and if the ultrasonic sensor is off.

(d) Extend the network from Figure 2(a) so that it can represent the following statements:
   (i) the alarm can be activated by pets;
   (ii) when the alarm goes on, an SMS notification is sent automatically via the Internet.

Solution:

(a) The joint probability of the network in Figure 2 (a) is

$$\mathbf{P}(B, U, A, N) = \mathbf{P}(B)\mathbf{P}(U)\mathbf{P}(A|B, U)\mathbf{P}(N|A).$$

The conditional probability table is

B	U	$P(A=0\|B,U)$	$P(A=1\|B,U)$
0	0	$1-\varepsilon$	ε
0	1	$1-2\varepsilon$	2ε
1	0	$1-x$	x
1	1	$1-y$	y

(b) From the Bayes theorem we have

$$\begin{aligned} P(B=1|A=1) &= \frac{P(A=1|B=1)P(B=1)}{P(A=1)} \\ &= \frac{P(A=1|B=1)P(B=1)}{P(A=1|B=0)P(B=0)+P(A=1|B=1)P(B=1)}. \end{aligned}$$

Using the conditional probability table from (a), we obtain

$$\begin{aligned} P(A=1|B=1) &= P(A=1|B=1,U=0)P(U=0)+P(A=1|B=1,U=1)P(U=1) \\ &= x(1-u)+yu \end{aligned}$$

and

$$\begin{aligned} P(A=1|B=0) &= P(A=1|B=0,U=0)P(U=0)+P(A=1|B=0,U=1)P(U=1) \\ &= \varepsilon(1-u)+2\varepsilon u. \end{aligned}$$

With $P(B=1)=b$ we can express

$$P(B=1|A=1) = \frac{b[x(1-u)+yu]}{(1-b)[\varepsilon(1-u)+2\varepsilon u]+b[x(1-u)+yu]}.$$

(c) We need to express $P(B = 1|N = 1, U = 0)$ by using the given network structure from (a): $\mathbf{P}(B, U, A, N) = \mathbf{P}(B)\mathbf{P}(U)\mathbf{P}(A|B, U)\mathbf{P}(N|A)$ and

$$\begin{aligned} P(N = 1|A = 0) &= 0 \\ P(N = 1|A = 1) &= n. \end{aligned}$$

We can write

$$\begin{aligned} P(B = 1|N = 1, U = 0) &= \alpha P(B = 1, N = 1, U = 0) \\ &= \alpha[P(B = 1, N = 1, U = 0, A = 0) \\ &\qquad + P(B = 1, N = 1, U = 0, A = 1)] \\ &= \alpha[b(1 - u)(1 - x) \cdot 0 + b(1 - u)x \cdot n] \\ &= \alpha b(1 - u)xn \end{aligned}$$

and also

$$\begin{aligned} P(B = 0|N = 1, U = 0) &= \alpha P(B = 0, N = 1, U = 0) \\ &= \alpha[P(B = 0, N = 1, U = 0, A = 0) \\ &\qquad + P(B = 0, N = 1, U = 0, A = 1)] \\ &= \alpha[(1 - b)(1 - u)(1 - \varepsilon) \cdot 0 + (1 - b)(1 - u)\varepsilon \cdot n] \\ &= \alpha(1 - b)(1 - u)\varepsilon n. \end{aligned}$$

From $P(B = 1|N = 1, U = 0) + P(B = 0|N = 1, U = 0) = 1$ we have that:

$$\alpha b(1 - u)xn + \alpha(1 - b)(1 - u)\varepsilon n = \alpha(1 - u)n[bx + (1 - b)\varepsilon] = 1$$

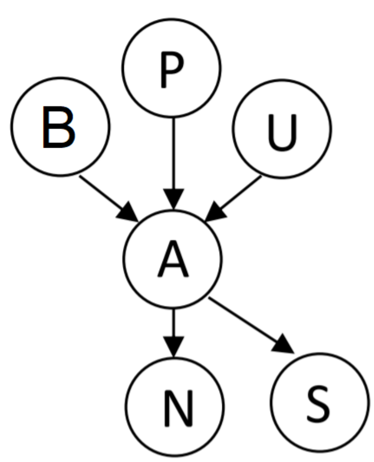
which yields

$$\alpha = \frac{1}{(1 - u)n[bx + (1 - b)\varepsilon]}$$

and we finally obtain

$$P(B = 1|N = 1, U = 0) = \frac{bx}{bx + (1 - b)\varepsilon}.$$

(d) The network should be extended with two additional nodes P (pets activating alarm) and S (SMS notification sent when alarm goes on) as follows:



3. PacLabs has just created a new type of mini power pellet that is small enough for Pacman to carry around with him when he's running around mazes. Unfortunately, these mini-pellets don't guarantee that Pacman will win all his fights with ghosts, and they look just like the regular dots Pacman carried around to snack on.

   Pacman just ate a snack (P), which was either a mini-pellet (p), or a regular dot ($\neg p$), and is about to get into a fight (W), which he can win (w) or lose ($\neg w$). Both these variables are unknown, but fortunately, Pacman is a master of probability. He knows that his bag of snacks has 5 mini-pellets and 15 regular dots. He also knows that if he ate a mini-pellet, he has a 70% chance of winning, but if he ate a regular dot, he only has a 20% chance.

   (a) What is $P(w)$, the marginal probability that Pacman will win?

   (b) Pacman won! Hooray! What is the conditional probability $P(p \mid w)$ that the food he ate was a mini-pellet, given that he won?

   (c) Pacman can make better probability estimates if he takes more information into account. First, Pacman's breath, B, can be bad (b) or fresh ($\neg b$). Second, there are two types of ghost (M): mean (m) and nice ($\neg m$). Pacman has encoded his knowledge about the situation in the following probability distribution

$$\mathbf{P}(M, P, B, W) = \mathbf{P}(M)\mathbf{P}(P)\mathbf{P}(W \mid M, P)\mathbf{P}(B \mid P).$$

   Based on the given probability distribution, in the box below draw the corresponding Bayesian network.

$\mathbf{P}(M)$	
m	0.5
$\neg m$	0.5

$\mathbf{P}(W \mid M, P)$			
m	p	w	0.60
m	p	$\neg w$	0.40
m	$\neg p$	w	0.10
m	$\neg p$	$\neg w$	0.90
$\neg m$	p	w	0.80
$\neg m$	p	$\neg w$	0.20
$\neg m$	$\neg p$	w	0.30
$\neg m$	$\neg p$	$\neg w$	0.70

$\mathbf{P}(P)$	
p	0.25
$\neg p$	0.75

$\mathbf{P}(B \mid P)$		
p	b	0.80
p	$\neg b$	0.20
$\neg p$	b	0.40
$\neg p$	$\neg b$	0.60

   (d) Just based on the structure, which of the following are guaranteed to be true? Explain your answers.

   (1) $W \perp\!\!\!\perp B$

   (2) $W \perp\!\!\!\perp B \mid P$

   (3) $M \perp\!\!\!\perp B$

   (4) $M \perp\!\!\!\perp B \mid P$

Solution:

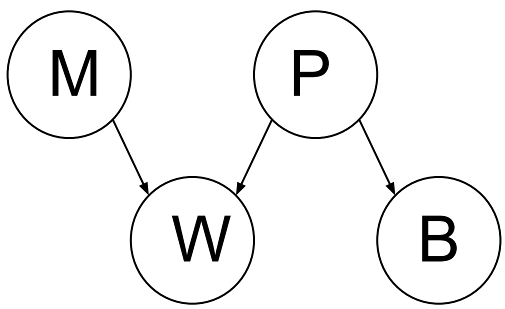
(a) After summing out the hidden variable P we obtain

$$\begin{aligned} P(w) &= P(w,p) + P(w,\neg p) = P(w \mid p)P(p) + P(w \mid \neg p)P(\neg p) \\ &= \frac{7}{10} \cdot \frac{5}{20} + \frac{2}{10} \cdot \frac{15}{20} = \frac{13}{40} = 0.325. \end{aligned}$$

(b) By using the Bayes' formula we get that

$$P(p \mid w) = \frac{P(w \mid p)P(p)}{P(w)} = \frac{\frac{7}{10} \cdot \frac{1}{4}}{\frac{13}{40}} = \frac{7}{13} \approx 0.538.$$

(c) From the probability distribution we infer that M and P have no parents. Similarly we see that W has two parents, namely M and P. Finally, B has one parent P.



(d) (2), (3) and (4) can be guaranteed to be true and (1) not (the divergent node P without evidence does not block the path between W and B).