Bayesian Networks 2016–2017

Tutorial I – Solutions

Peter Lucas
LIACS, Leiden University

Exercise 1

a. Draw a Venn diagram with x and y and intersection $x \wedge y$. You will see that the area $x \wedge y$ is part of both x and y and thus counted twice if you compute $P(x \vee y)$ in terms of $P(x)$ and $P(y)$.

b. Proof:

$$\begin{aligned} x \vee y &\equiv (x \vee \neg x) \wedge (x \vee y) \\ &\equiv x \vee (\neg x \wedge y) \\ &\equiv (x \wedge \neg y) \vee (x \wedge y) \vee (\neg x \wedge y) \end{aligned}$$

these Boolean expressions are mutually exclusive.
$\Rightarrow P(x \vee y) = P(x \wedge \neg y) + P(x \wedge y) + P(\neg x \wedge y)$ (1). Using marginalisation: $P(x) = P(x \wedge y) + P(x \wedge \neg y)$ (2) and $P(y) = P(y \wedge x) + P(y \wedge \neg x)$ (3). Using (1), (2) and (3) $P(x \vee y) = P(x) + P(\neg x \wedge y) = P(x) + P(y) - P(x \wedge y)$.

Exercise 2

a. $P(a) = P(a, b) + P(a, \neg b) = 0.3 + 0.4 = 0.7$; $P(b) = P(a, b) + P(\neg a, b) = 0.3 + 0.2 = 0.5$.

b. $P(a \mid b) = \frac{P(a,b)}{P(b)} = 0.3/0.5 = 0.6$.

c. $P(b \mid a) = \frac{P(a|b)P(b)}{P(a)} = \frac{0.6 \cdot 0.5}{0.7} = 3/7$.

Exercise 3

a. Proof:

$$P(X \mid Y) = P(X, Y)/P(Y)$$

and

$$P(Y \mid X) = P(Y, X)/P(X)$$

as $P(X, Y) = P(Y, X)$ (commutativity of the conjunction) we have that

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)}$$

b. If you have the Bayesian network $X \to Y$ with parameter $P(Y \mid X)$ associated with the arc then you can change it into $X \leftarrow Y$ by computing the parameters $P(X \mid Y)$ using Bayes' rule from the old Bayesian network's parameters.

c. Computation would be in terms of $P(A_1, \ldots, A_n \mid B)$ and each of the probability distributions is exponential in the number of variable A_i. Solution is to take into account independence information.

Exercise 4

a. $P(x_2 \vee \neg x_3 \mid x_1 \wedge x_4) = P(x_2 \mid x_1 \wedge x_4) + P(\neg x_3 \mid x_1 \wedge x_4) - P(x_2 \wedge \neg x_3 \mid x_1 \wedge x_4)$. Each of these terms can be computed easily using marginalisation and conditional probabilities. For example, $P(x_2 \mid x_1 \wedge x_4) = P(x_2, x_1, x_4)/P(x_1, x_4)$ and $P(x_2, x_1, x_4) = P(x_1, x_2, x_3, x_4) + P(x_1, x_2, \neg x_3, x_4) = 0.1 + 0.03 = 0.13$, etc.

b. Many possible answers; one of them: $P(X_1, X_2, X_3, X_4) = P(X_1 \mid X_2, X_3, X_4)P(X_2 \mid X_3, X_4)P(X_3, X_4)$.

Computation of these factors, for example: $P(x_3, x_4) = P(x_1, x_2, x_3, x_4) + P(\neg x_1, x_2, x_3, x_4) + P(x_1, \neg x_2, x_3, x_4) + P(\neg x_1, \neg x_2, x_3, x_4)$; see problem for probabilities.

Exercise 5

a. $P(x \mid y, z) = 0.3$, $P(\neg x \mid y, z) = 0.7$, $P(x \mid \neg y, z) = 0.3$, $P(\neg x \mid \neg y, z) = 0.7$, $P(x \mid y, \neg z) = 0.4$, $P(x \mid \neg y, \neg z) = 0.4$, etc.

b. What is to be proven is that $P(X \mid Y, Z) = P(X \mid Z)(1) \Rightarrow P(Y \mid X, Z) = P(Y \mid Z)$ (2) for any value of X, Y, Z. Proof:

$$\begin{aligned} P(Y \mid X, Z) &= P(X, Z \mid Y)P(Y)/P(X, Z) \\ &= (P(X \mid Y, Z)P(Z \mid Y)P(Y)) \,/\, (P(X \mid Z)P(Z)) \\ &\stackrel{(1)}{=} (P(X \mid Z)P(Z, Y))/(P(X \mid Z)P(Z)) \\ &= P(Z, Y)/P(Z) \\ &= P(Y \mid Z) \end{aligned}$$

So, $P(Y \mid X, Z) = P(Y \mid Z)$ if $X \perp\!\!\!\perp_P Y \mid Z$.

Exercise 6

a. Proof:

$$\begin{aligned} P(X \mid Y, Z) &= P(X, Y \mid Z)/P(Y \mid Z) \\ \Leftrightarrow P(X, Y \mid Z) &= P(X \mid Y, Z)P(Y \mid Z) \\ & \quad \text{Using } P(X \mid Y, Z) = P(X \mid Z) \text{, yields} \\ P(X, Y \mid Z) &= P(X \mid Z)P(Y \mid Z) \end{aligned}$$

b. In Bayesian networks we often compute probabilities $P(X \mid Y)$, where $P(X \mid \cdots)$ is obtained by marginalisation of a conditional probability.

Exercise 7

a. We have a Bayesian network with variable X with associated probability distribution $P(X)$ that is computed by $f(x; p)$ and $f(x; p, n)$ respectively.

b. We have a Bayesian network with variable X with one parent variable Θ that corresponds to the parameter p for the Bernoulli case and with a second parent variable N for the binomial case. Such Bayesian networks with parameters explicitly represented are called *augmented* Bayesian networks. The prior density on Θ might be a beta distribution, whereas for N it might be a Poisson distribution.

Exercise 8

We have by Bayes' rule:

$$\begin{aligned} P(\mathit{flu} \mid \mathit{temp} > 37.5) &= P(\mathit{temp} > 37.5 \mid \mathit{flu})P(\mathit{flu})/P(\mathit{temp} > 37.5) \\ &= (1 - P(\mathit{temp} \leq 37.5 \mid \mathit{flu}))P(\mathit{flu})/P(\mathit{temp} > 37.5) \end{aligned}$$

$P(\mathit{temp} > 37.5)$ can be computed by marginalisation and conditioning ($P(\mathit{temp} > 37.5) = P(\mathit{temp} > 37.5 \mid \mathit{flu})P(\mathit{flu}) + P(\mathit{temp} > 37.5 \mid \mathit{pneu})P(\mathit{pneu})$), but alternatively one can make use of the trick that $P(\mathit{flu} \mid \mathit{temp} > 37.5)$ and $P(\mathit{pneu} \mid \mathit{temp} > 37.5)$ add up to 1 and have the same denominator $P(\mathit{temp} > 37.5)$ in their Bayes' rule reformulation.

Exercise 9

Consider Figure 1, which displays a Bayesian network in which the three vertices A, B and C interact to cause effect D through a noisy-AND. The intermediate variables I_A, I_B and I_C are note indicated in the figure, but it is assumed that for computational purposes these intermediate variables are implicitly present. The following probabilities have been specified by the designer of the Bayesian network:

$$\begin{array}{ll} P(i_A \mid a) = 0.7 & P(i_A \mid \neg a) = 0.9 \\ P(i_B \mid b) = 0.4 & P(i_B \mid \neg b) = 0.8 \\ P(i_C \mid c) = 0.3 & P(i_C \mid \neg c) = 0.3 \end{array}$$

$$\begin{array}{ll} P(a) = 0.4 & P(b) = 0.7 \\ P(c) = 0.8 & \end{array}$$

$$\begin{array}{ll} P(e \mid d) = 0.2 & P(e \mid \neg d) = 0.6 \end{array}$$

a. Compute $P^*(e) = P(e \mid a, b, c)$, i.e. the marginal probability of e given that $A = B = C = \mathit{true}$.

b. Compute $P^*(e) = P(e \mid a, b)$.

Solutions

a.
$$\begin{aligned} P(d \mid a, b, c) &= \textstyle\sum_{I_A, I_B, I_C} P(d \mid I_A, I_B, I_C)P(I_A \mid a)P(I_B \mid b)P(I_C \mid c) \\ &= \textstyle\sum_{I_A \wedge I_B \wedge I_C = \mathit{true}} P(I_A \mid a)P(I_B \mid b)P(I_C \mid c) \\ &\quad \text{(because we model an AND)} \\ &= 0.7 \cdot 0.4 \cdot 0.3 = 0.084 \end{aligned}$$

Now we have that

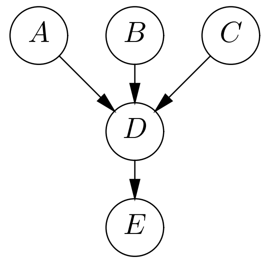

Figure 1: Bayesian network: noisy-AND.

$$\begin{aligned} P(e \mid a, b, c,) &= \textstyle\sum_D P(e, D \mid a, b, c,) \\ &= \textstyle\sum_D P(e \mid D) P(D \mid a, b, c,) \text{ (see above for } P(d \mid a, b, c)\text{)} \\ &= 0.2 \cdot 0.084 + 0.6 \cdot 0.916 = 0.5664 \end{aligned}$$

b.

$$\begin{aligned} P(e \mid a, b) &= \textstyle\sum_{C,D} P(e \mid C, D, a, b) P(C, D \mid a, b) \\ &= \textstyle\sum_{C,D} P(e \mid D) P(D \mid a, b, C) P(C \mid a, b) \\ &= \textstyle\sum_{C,D} P(e \mid D) P(D \mid a, b, C) P(C) \text{ (because } A \text{ and } B \text{ are independent of } C\text{)} \end{aligned}$$

We need to compute $P(d \mid a, b, c)$ (see above) and $P(d \mid a, b, \neg c)$. It appears that in this case we have that $P(d \mid a, b, c) = P(d \mid a, b, \neg c)$ (because in this case $P(i_C \mid c) = P(i_C \mid \neg c)$). If you fill in the equation you also get in this case that $P(e \mid a, b) = 0.5664$.

Exercise 10

a. See the network of lecture 2, page 31.

b. See lecture 2, pages 32-33.

c. $P^*(v_2) = P(v_2 \mid v_2) = 1$, $P^*(\neg v_2) = P(\neg v_2 \mid v_2) = 0$, and

$$\begin{aligned} P^*(v_4) &= P(v_4 \mid v_2) \\ &= P(v_4 \mid v_3) P(v_3 \mid v_2) + P(v_4 \mid \neg v_3) P(\neg v_3 \mid v_2) \end{aligned}$$

where

$$\begin{aligned} P(V_3 \mid v_2) &= P(V_3 \mid v_1, v_2) P(v_1 \mid v_2) + P(V_3 \mid \neg v_1, v_2) P(\neg v_1 \mid v_2) \\ &= P(V_3 \mid v_1, v_2) P(v_1) + P(V_3 \mid \neg v_1, v_2) P(\neg v_1) \end{aligned}$$

because $V_1 \perp\!\!\!\perp_P V_2 \mid \varnothing$. The rest is just filling in numbers. $P^*(v_1) = P(v_1)$.

d. This is easy as (Bayes' rule again):

$$\begin{aligned} P^*(v_2) &= P(v_2 \mid v_4) \\ &= P(v_4 \mid v_2) P(v_2) / P(v_4) \\ &= \sum_{V_3} P(v_4, V_3 \mid v_2) P(v_2) / P(v_4) \\ &= \sum_{V_3} P(v_4 \mid V_3) P(V_3 \mid v_2) P(v_2) / P(v_4) \end{aligned}$$

$$= \sum_{V_3} P(v_4 \mid V_3) \sum_{V_1} P(V_3 \mid V_1, v_2) P(V_1) P(v_2) / P(v_4)$$

$P(v_4)$ is computed in slides 32-33 of Lecture 2. The rest of the probability distributions can be looked up in the Bayesian network.

e. Use the factorisation of the Bayesian network

$$P(v_1, \neg v_2, v_3, v_4) = P(v_4 \mid v_3) P(v_3 \mid v_1, \neg v_2) P(v_1) P(\neg v_2)$$

Exercise 11

a. Again Bayes' rule:

$$P(\neg v_5 \mid v_3) = P(v_3 \mid \neg v_5) P(v_5) / P(v_3)$$

Hence, we only have to compute $P(v_3 \mid \neg v_5) = \sum_{V_2, V_4} P(v_3, V_2, V_4 \mid \neg v_5) = \cdots$.

b. If the inference algorithm does not take this into account, the probabilistic information may cycle for ever.

Exercise 12

a. $V_1 \perp\!\!\!\perp_G V_4 \mid V_3$
$V_2 \perp\!\!\!\perp_G V_4 \mid V_3$
$\{V_1, V_2\} \perp\!\!\!\perp_G V_4 \mid V_3$, plus the independence axioms applied (such as symmetry, etc).

b. $V_1 \perp\!\!\!\perp_G V_2 \mid \varnothing$, plus the independence axioms.

c. $V_1 \not\perp\!\!\!\perp_G V_2 \mid V_3$.
$V_1 \not\perp\!\!\!\perp_G V_2 \mid V_4$
$V_1 \not\perp\!\!\!\perp_G V_2 \mid \{V_3, V_4\}$ $V_3 \not\perp\!\!\!\perp_G V_4 \mid W$
$V_2 \not\perp\!\!\!\perp_G V_1 \mid W$
$V_3 \not\perp\!\!\!\perp_G V_2 \mid W$, any proper W,
$V_3 \not\perp\!\!\!\perp_G \{V_1, V_2\} \mid W$,
$V_4 \not\perp\!\!\!\perp_G V_2 \mid U$, any U with $V_3 \notin U$
$V_4 \not\perp\!\!\!\perp_G V_1 \mid U$, any U with $V_3 \notin U$, and some combinations.