Problems for Chapter 14*

Anca Ralescu

7/25/2021

Problem 14.1 We have a bag with three biased coins a, b, and c with probabilities of coming up heads of 0.2, 0.6, and 0.8 respectively. Each coin is equally likely to be drawn from the bag. A coin is drawn and then it is flipped three times, yielding the outcomes X_1, X_2, X_3.

(a) Draw the Bayes network corresponding to this setup and define the necessary CPTs.

(b) Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Solution **(a)** Let C be a random variable denoting which coin $\{a, b, c\}$ we draw. Let C be the root of the network and $X1, X2$, and $X3$ as children. The CPT for C is shown in Table :

C	P(C)
a	1/3
b	1/3
c	1/3

The CPT for Xi, $i = 1, 2, 3$ given C are the same, shown in table

C	X1	P(C)
a	heads	0.2
b	heads	0.6
c	heads	0.8

(b) The coin most likely to have been drawn from the bag given this sequence is the value of C with greatest posterior probability $P(C|2 \text{ heads, } 1 \text{ tails})$. Now,

$$\begin{aligned} P(C|2 \text{ heads, } 1 \text{ tails}) = & \frac{P(2 \text{ heads, } 1 \text{ tails} \mid \text{C})\text{P(C)})}{P(2 \text{ heads, } 1 \text{ tails})} \\ & \propto P(2 \text{ heads, } 1 \text{ tails} \mid C)P(C) \\ & \propto P(2 \text{ heads, } 1 \text{ tails} \mid C) \end{aligned}$$

where in the second line we observe that the constant of proportionality

$$\frac{1}{P(2 \text{ heads, } 1 \text{ tails})}$$

is independent of C, and in the last we observe that $P(C)$ is also independent of the value of C since it is, by hypothesis, equal to 1/3.

*I am also including problem 9.24 of the textbook.

From the Bayesian network we can see that $X1, X2$, and $X3$ are **conditionally independent given** C, so for example

$$P(X1 = \text{tails},\ X2 = \text{heads},\ X3 = \text{heads}|C = a) =$$
$$P(X1 = \text{tails} \mid C = a)P(X2 = \text{heads} \mid C = a)P(X3 = \text{heads} \mid C = a)$$
$$= 0.8 \text{ x } 0.2 \text{ x } 0.2 = 0.032$$

Since the CPTs for each coin are the same, we would get the same probability above for any ordering of 2 heads and 1 tails.

Since there are three such orderings, we have

$$P(2 \text{ heads, } 1\text{tails}|C = a) = 3 \text{ x } 0.032 = 0.096$$

Similar calculations to the above find that

$$P(2\text{heads, } 1\text{tails}|C = b) = 0.432$$

$$P(2\text{heads, } 1\text{tails}|C = c) = 0.384$$

showing that coin b is most likely to have been drawn. Alternatively, one could directly compute the value of $P(C|2 \text{ heads, } 1 \text{ tails})$.

Problem 14.2 Equation (14.1) on page 513 defines the joint distribution represented by a Bayesian network in terms of the parameters $\theta(X_i|Parents(X_i))$. This exercise asks you to derive the equivalence between the parameters and the conditional probabilities $P(X_i|Parents(X_i))$ from this definition.

a. Consider a simple network $X \to Y \to Z$ with three Boolean variables. Use Equations (13.3) and (13.6) (pages 485 and 492) to express the conditional probability $P(z|y)$ as the ratio of two sums, each over entries in the joint distribution $P(X, Y, Z)$.

b. Now use Equation (14.1) to write this expression in terms of the network parameters $\theta(X), \theta(Y|X)$, and $\theta(Z|Y)$.

c. Next, expand out the summations in your expression from part (b), writing out explicitly the terms for the true and false values of each summed variable. Assuming that all network parameters satisfy the constraint $\sum_{x_i} \theta(x_i|parents(X_i)) = 1$, show that the resulting expression reduces to $\theta(x|y)$.

d. Generalize this derivation to show that $\theta(X_i|Parents(X_i)) = P(X_i|Parents(Xi))$ for **any** Bayesian network.

SOLUTION Equations (13.3) and (13.6)

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \quad (13.3)$$

$$\text{marginalization: } \mathbf{P}(\mathbf{Y}) = \sum_{z \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, z) \quad (13.6)$$

a. By Equations (13.3) and (13.6), we have

$$P(z|y) = \frac{P(y, z)}{P(y)} = \frac{\sum_x P(x, y, z)}{\sum_{x,x'} P(x, y, z)}.$$

b. By Equation (14.1), this can be written as

$$P(z|y) = \frac{\sum_x \theta(x)\theta(y|x)\theta(z|y)}{\sum_{x,z} \theta(x)\theta(y|x)\theta(z|y)}.$$

c. The expanding-out step makes it a little easier to see how to simplify the expressions. Expanding out the sums, collecting terms, using the sum-to-1 property of the parameters, and finally canceling, we have

$$\begin{aligned}
P(z|y) &= \frac{\theta(x)\theta(y|x)\theta(z|y)+\theta(\neg x)\theta(y|\neg x)\theta(z|y)}{\theta(x)\theta(y|x)\theta(z|y)+\theta(x)\theta(y|x)\theta(\neg z|y)+\theta(\neg x)\theta(y|\neg x)\theta(z|y)+\theta(\neg x)\theta(y|\neg x)\theta(\neg z|y)} \\
&= \frac{\theta(z|y)[\theta(x)\theta(y|x)+\theta(\neg x)\theta(y|\neg x)]}{[\theta(x)\theta(y|x)+\theta(\neg x)\theta(y|\neg x)][\theta(z|y)+\theta(\neg z|y)]} \\
&= \frac{\theta(z|y)[\theta(x)\theta(y|x)+\theta(\neg x)\theta(y|\neg x)]}{[\theta(x)\theta(y|x)+\theta(\neg x)\theta(y|\neg x)]} \\
&= \theta(z|y).
\end{aligned}$$

Instead, working on the summations directly, the key step is moving the sum over z inwards:

$$\begin{aligned}
P(z|y) &= \frac{\theta(z|y)\sum_x \theta(x)\theta(y|x)}{\sum_x \theta(x)\theta(y|x)\sum_z \theta(z|y)} \\
&= \frac{\theta(z|y)\sum_x \theta(x)\theta(y|x)}{\sum_x \theta(x)\theta(y|x)} \\
&= \theta(z|y).
\end{aligned}$$

d. The general case is more difficult (I omit it)

Problem 14.3 The operation of arc reversal in a Bayesian network allows us to change the direction of an arc $X \to Y$ while preserving the joint probability distribution that the network represents (Shachter, 1986). Arc reversal may require introducing new arcs: all the parents of X also become parents of Y , and all parents of Y also become parents of X.

a. Assume that X and Y start with m and n parents, respectively, and that all variables have k values. By calculating the change in size for the CPTs of X and Y , show that the total number of parameters in the network cannot decrease during arc reversal. (Hint: the parents of X and Y need not be disjoint.)

b. Under what circumstances can the total number remain constant?

c. Let the parents of X be $\mathbf{U} \cup \mathbf{V}$ and the parents of Y be $\mathbf{V} \cup \mathbf{W}$, where $\mathbf{U}$ and $\mathbf{W}$ are disjoint. The formulas for the new CPTs after arc reversal are as follows:

$$P(Y|\mathbf{U}, \mathbf{V}, \mathbf{W}) = \sum_x P(Y|\mathbf{V}, \mathbf{W}, x)P(x|\mathbf{U}, \mathbf{V})$$

$$P(X|\mathbf{U}, \mathbf{V}, \mathbf{W}, Y) = P(Y|X, \mathbf{V}, \mathbf{W})P(X|\mathbf{U}, \mathbf{V})/P(Y|\mathbf{U}, \mathbf{V}, \mathbf{W}).$$

Prove that the new network expresses the same joint distribution over all variables as the original network.

SOLUTION 14.3

a. Suppose that X and Y share l parents. After the reversal Y will gain m-l new parents, the ml original parents of X that it does not share with Y , and loses one parent: X. After the reversal X will gain $n - l$ new parents, the $n - l - 1$ original parents of Y that it does not share with X and isnt X itself, and plus Y. So, after the reversal Y will have $n + (m - l - 1) = m + (n - l - 1)$ parents, and X will have $m + (n - l) = n + (m - l)$ parents.

$m - l \geq 0$, since this is the number of original parents of X not shared with Y , and that $n - l - 1 \geq 0$, since this is the number of original parents of Y not shared with X and not equal to X.

This shows the number of parameters can only increase: before we had $km + kn$, after we have $km + (n - l - 1) + kn + (m - l)$.

Check the counting above, if are reversing a single arc without any extra parents, we have $l = 0, m = 0$, and $n = 1$; the previous formulas say we will have $m = 0$ and $n = 1$ afterwards, which is correct.

b. For the number of parameters to remain constant, assuming that $k > 1$, requires by our previous calculation that $m - l = 0$ and $n - l - 1 = 0$. This holds exactly when X and Y share all their parents originally (except Y also has X as a parent).

c. Tedious (omitted).

Problem 14.4 The network in Figure 14.2

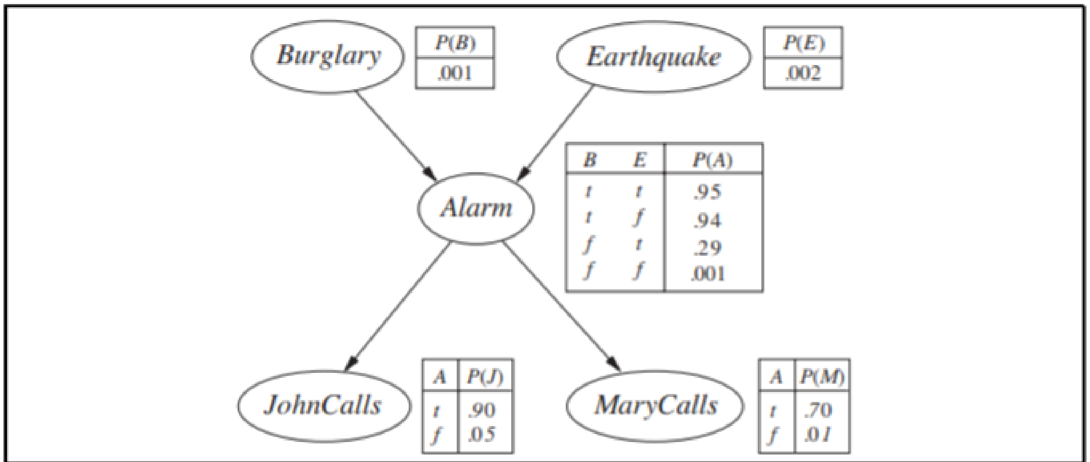

Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

CPT are as follows:

$$P(B = yes) = 0.001; P(Earthquake = yes) = 0.002$$

CPT for Alarm : CPT for JohnCalls: Table CPT for MaryCalls: Table

B	E	P(A)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	P(J)
t	0.90
f	0.05

A	P(M)
t	0.70
f	0.01

Solution

(a) Yes. Numerically one can compute that $P(B, E) = P(B)P(E)$. Topologically B and E are d-separated by A.

(b) We check whether $P(B,E|a) = P(B|a)P(E|a)$. First computing $P(B,E|a)$:

$$\begin{aligned} P(B,E|a) &= \alpha P(a|B,E)P(B,E) \\ &= \alpha \begin{cases} .95 \times 0.001 \times 0.002 & \text{if} \quad B = b \text{ and } E = e \\ .94 \times 0.001 \times 0.998 & \text{if} \quad B = b \text{ and } E = \neg e \\ .29 \times 0.999 \times 0.002 & \text{if} \quad B = \neg b \text{ and } E = e \\ .001 \times 0.999 \times 0.998 & \text{if} \quad B = \neg b \text{ and } E = \neg e \end{cases} \\ &= \alpha \begin{cases} 0.0008 & \text{if} \quad B = b \text{ and } E = e \\ 0.3728 & \text{if} \quad B = b \text{ and } E = \neg e \\ 0.2303 & \text{if} \quad B = \neg b \text{ and } E = e \\ 0.3962 & \text{if} \quad B = \neg b \text{ and } E = \neg e \end{cases} \end{aligned}$$

where α is a normalization constant. Checking whether

$$P(b,e|a) = P(b|a)P(e|a)$$

we find

$$P(b,e|a) = 0.0008 \neq 0.0863 = 0.3736 \times 0.2311 = P(b|a)P(e|a)$$

showing that B and E are not conditionally independent given A.

Problem 9.24 in the textbook Domain: natural numbers, i.e,
$\mathbf{N} = \{0, 1, 2, \ldots, \infty\}$

$$(A)\ \ \forall x\ \exists y\ (x \geq y)$$

$$(B)\ \ \exists y\ \forall x (x \geq y)$$

(a) Translate (A) and (B) in English.

(b) Is (A) true?

(c) Is (B) true?

(d) Does (A) logically entail (B)?

(e) Does (B) logically entail (A)?

(f) Using resolution try to prove (d), that is, that (A) follows from (B). Do this even if you think that it is not true and see where the proof gets stuck.

(g) Using resolution try to prove (e), that is, that (B) follows from (A). Do this even if you think that it is not true and see where the proof gets stuck.

Solution

(a) $(A)\ \ \forall x\ \exists y\ (x \geq y)$ translates to "*For every natural number there is some other natural number that is smaller than or equal to it.*"

$(B)\ \ \exists y\ \forall x (x \geq y)$ translates to "*There is a particular natural number that is smaller than or equal to any natural number.*"

(b) Yes, (A) is true: one can always pick the number itself for the "*some other*" number.

(c) Yes, (B) is true: one can pick 0 for the "*particular natural number.*"

(d) No, (A) does not logically entail (B).

(e) Yes, (B) logically entails (A).

(f) We want to try to prove via resolution that (A) entails (B). To do this, we set our knowledge base to consist of (A) and the negation of (B), which we will call (¬B), and try to derive a contradiction. First we have to convert (A) and (¬B) to canonical form.

Let us use the predicate *ge* for ≥ so that (A) and (B) become

$$(A)\ \ \forall x\ \exists y\ ge(x, y)$$

and

$$(B)\ \ \exists y\ \forall x ge(x, y)$$

To figure out (¬B) we need to move ¬ in, past the two quantifiers.

Both sentences will use Skolem functions:

$$(A)\ \ ge(x, F_A(x))$$

and

$$(\neg B) \neg ge(F_B(y), y)$$

Now we can try to resolve these two together, using the rules for *unification*. We look for a substitution with will make $ge(x, F_A(x))$ and $ge(F_B(y), y)$ equal: It looks like the substitution should be $\{x/F_B(y), y/F_A(x)\}$, but that is equivalent to $\{x/F_B(y), y/F_A(F_B(y))\}$, which fails because y would have to be bound to an expression containing itself. So this resolution step fails, and since there are no other resolution steps to try, concludethat (B) does not follow from (A).

(g) To prove that (B) entails (A), we start with a knowledge base containing (B) and the negation of (A), which we will call (¬A). Use *ge* predicate as above to obtain:

$$(\neg A): \ \ \neg ge(X_0, y), \ \ \text{where } X_0 \text{ is a constant from the domain; simple drop of } \exists$$

and

$$(B): \ \ ge(x, Y_0), \ \ \text{where } Y_0 \text{ is a constant from the domain; simple drop of } \exists$$

This time the resolution goes through, with the substitution $\{x/X_0, y/Y_0\}$, thereby yielding False, and therefore proving that (B) entails (A).