
Exercises 3

Directed graphical models

3.1 Bayesian network basics

Exercise 3.1 Number of parameters

Consider a probabilistic model involving three binary random variables A , B , and C . In the general case, there are $2^3 = 8$ possible outcomes for the three binary random variables and we need up to 7 distinct probabilities to specify the joint distribution of the random variables,

$$\begin{aligned} p(A, B, C) = p_1, \quad p(\bar{A}, B, C) = p_2, \quad p(A, \bar{B}, C) = p_3, \quad p(A, B, \bar{C}) = p_4, \\ p(A, \bar{B}, \bar{C}) = p_5, \quad p(\bar{A}, B, \bar{C}) = p_6, \quad p(\bar{A}, \bar{B}, C) = p_7, \end{aligned} \quad (17)$$

as the eighth comes from the constraint that the probabilities need to sum to 1: $p(\bar{A}, \bar{B}, \bar{C}) = 1 - p_1 - p_2 - p_3 - p_4 - p_5 - p_6 - p_7$.

- Suppose that the joint distribution is factorized according to the Bayesian network in Figure 3.1 (left). How many parameters are needed to specify the joint distribution? Express these parameters in terms of the parameters in (17).
- Suppose that the joint distribution is factorized according to the Bayesian network in Figure 3.1 (right). How many parameters are needed to specify the joint distribution? Express these parameters in terms of the parameters in (17).



Figure 3.1: Bayesian networks in Exercise 3.1

Exercise 3.2 D-separation basic structures

Consider the three Bayesian networks in Figure 3.2.

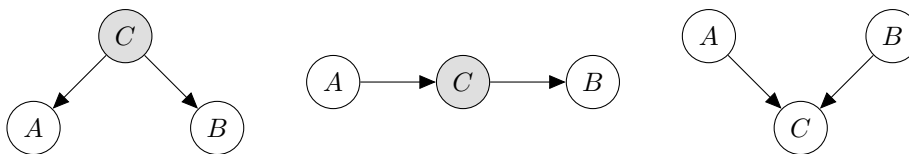


Figure 3.2: Bayesian networks in Exercise 3.2

Show that

Table 3.1: Joint distribution of the variables in Exercise 3.4

A	B	C	$p(A, B, C)$
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

- a) for the left network, $A \perp\!\!\!\perp B|C$;
- b) for the center network, $A \perp\!\!\!\perp B|C$;
- c) for the right network, $A \perp\!\!\!\perp B|\emptyset$.

Exercise 3.3 Marginalization (adapted from [5])

Consider the Bayesian network in Figure 3.3. Suppose that we marginalize out the random variable X . What does the resulting Bayesian network look like? Which extra edges need to be added?

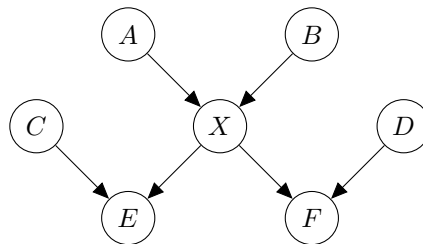


Figure 3.3: Bayesian network in Exercise 3.3.

Exercise 3.4 (adapted from [2])

Consider the three binary random variables A , B , and C having the joint distribution in table 3.1. Show, by explicit computation, that A and B are marginally dependent (i.e., $p(A, B) \neq p(A)p(B)$) and that they become independent when conditioned on C (i.e., $p(A, B|C) = p(A|C)p(B|C)$).

3.2 Modeling

Exercise 3.5 The Krakozhia problem (adapted from [4] and [1])

You are tasked with developing a clinical model for the *ChestPain Inc.* clinic to diagnose tuberculosis, lung cancer, and bronchitis. Smoking is a known risk factor in bronchitis and lung cancer and, recently, there was an outbreak of tuberculosis in Krakozhia, so prolonged stays in the country increase the risk of infection. Tuberculosis and lung cancer may cause lung damage. Lung damage and bronchitis are both causes of *dyspnoea*, or shortness of breath.

Express this information as a Bayesian network, and use that for determining whether the following statements are true or not:

- a) *tuberculosis* $\perp\!\!\!\perp$ *smoking* | *dyspnoea*
- b) *lung cancer* $\perp\!\!\!\perp$ *bronchitis* | *smoking*
- c) *stay in Krakozhia* $\perp\!\!\!\perp$ *smoking* | *lung cancer*
- d) *stay in Krakozhia* $\perp\!\!\!\perp$ *smoking* | *lung cancer, dyspnoea*

Exercise 3.6 Life of a builder (adapted from [1])

Exposure to asbestos (A) and smoking (S) have a synergistic effect on lung cancer (C) as described by

$$p(C, A, S) = p(C|A, S)p(A)p(S) \tag{18}$$

- a) Draw the graphical model representing (18).
- b) Conditioned on a cancer diagnosis, are S and A independent?
- c) Modify the model accounting for the fact that construction workers (W) are more likely to be exposed to asbestos.
- d) Given A , show that W and C are independent.

Exercise 3.7 D-separation

Consider the joint probability

$$p(A, B, C, D) = p(A|B)p(B|C)p(D|C)P(C). \tag{19}$$

- a) Draw the corresponding Bayesian network.
- b) Consider B and D ; which latent variables make them conditionally independent? Prove it explicitly.

Exercise 3.8 At the office (adapted from [3])

The manager Laura wants to track the working habits of the developer Marcus. Laura knows that Marcus spends 80% of the time at the office. In order not to disturb him while he works, she only checks for light at his window. Laura knows that Marcus sometimes works in the dark, and he keeps the light on 75% of the time he is at the office. He leaves the lights on when he leaves 2% of the times.

In addition to checking the window, Laura has access to the login information of the mainframe Marcus is working on. When he is at the office, Marcus is logged in 92% of the time. When he is at home, he logs in 15% of the time.

- a) Draw the Bayesian network representing the scenario just described.
- b) Compute Laura’s belief that Marcus’s light is on.
- c) Suppose that Laura sees that Marcus is logged into the mainframe. What effect does this observation have on her belief that Marcus’s light is on?

Exercise 3.9 An economic model (adapted from [1])

Consider the probabilistic model linking economy (E) to inflation (I) and price of oil (O) through observations of stock prices (S) and prices of oil futures (F). Suppose that all the quantities can take on two distinct values of high (h) and low (l), except for the stock price which can also be normal (n).

Table 3.2: Conditional probabilities of the model in Exercise 3.9

$p(E = l) = 0.2$	
$p(S = l O = l) = 0.9$	$p(S = n O = l) = 0.1$
$p(S = l O = h) = 0.1$	$p(S = n O = h) = 0.4$
$p(O = l E = l) = 0.9$	$p(O = l E = h) = 0.05$
$p(F = l I = l, E = l) = 0.9$	$p(F = l I = l, E = h) = 0.1$
$p(F = l I = h, E = l) = 0.1$	$p(F = l I = h, E = h) = 0.01$
$p(I = l O = l, E = l) = 0.9$	$p(I = l O = l, E = h) = 0.1$
$p(I = l O = h, E = l) = 0.1$	$p(I = l O = h, E = h) = 0.01$

- a) Using the conditional probabilities given Table 3.2 draw the Bayesian network corresponding to this model.
- b) Then, compute the probability that inflation (I) is high, given that the stock price (S) is normal and the futures price (F) is high.

Solutions 3

Directed graphical models

Solution to Exercise 3.1 a) To specify a conditional probability, we need two parameters (one for when the antecedent is true and one for when it is false),

$$p(B|A) = q_1, \quad p(B|\bar{A}) = q_2 \quad p(\bar{B}|A) = 1 - q_1, \quad p(\bar{B}|\bar{A}) = 1 - q_2.$$

and similarly

$$p(C|B) = q_3, \quad p(C|\bar{B}) = q_4 \quad p(\bar{C}|B) = 1 - q_3, \quad p(\bar{C}|\bar{B}) = 1 - q_4.$$

In addition, we need one Bernoulli probability $p(A) = q_5$; so in total we need 5 parameters.

Using the law of total probability, we have that

$$\begin{aligned} p(A, B) &= \sum_c p(A, B, c) = p(A, B, C) + p(A, B, \bar{C}) = p_1 + p_4 \\ p(\bar{A}, B) &= \sum_c p(\bar{A}, B, c) = p(\bar{A}, B, C) + p(\bar{A}, B, \bar{C}) = p_2 + p_6 \\ p(A, \bar{B}) &= \sum_c p(A, \bar{B}, c) = p(A, \bar{B}, C) + p(A, \bar{B}, \bar{C}) = p_3 + p_5 \end{aligned}$$

Then

$$q_5 = p(A) = \sum_b p(A, b) = p(A, B) + p(A, \bar{B}) = p_1 + p_3 + p_4 + p_5;$$

So

$$\begin{aligned} q_1 = p(B|A) &= \frac{p(A, B)}{p(A)} = \frac{p_1 + p_4}{p_1 + p_3 + p_4 + p_5}, \\ q_2 = p(B|\bar{A}) &= \frac{p(\bar{A}, B)}{p(\bar{A})} = \frac{p_2 + p_6}{1 - p_1 - p_3 - p_4 - p_5}. \end{aligned}$$

Similarly,

$$\begin{aligned} p(B, C) &= p(A, B, C) + p(\bar{A}, B, C) = p_1 + p_2 \\ p(B, \bar{C}) &= p(A, B, \bar{C}) + p(\bar{A}, B, \bar{C}) = p_4 + p_6 \\ p(\bar{B}, C) &= p(A, \bar{B}, C) + p(\bar{A}, \bar{B}, C) = p_3 + p_7 \end{aligned}$$

So

$$\begin{aligned} q_3 = p(C|B) &= \frac{p(B, C)}{p(B)} = \frac{p_1 + p_2}{p_1 + p_2 + p_4 + p_6}, \\ q_4 = p(C|\bar{B}) &= \frac{p(\bar{B}, C)}{p(\bar{B})} = \frac{p_3 + p_7}{1 - p_1 - p_2 - p_4 - p_6}. \end{aligned}$$

- b) To specify the conditional probability of $C|A, B$, we need four parameters (one for each combination of the antecedents):

$$p(C|A, B) = r_1, \quad p(C|\bar{A}, B) = r_2 \quad p(C|A, \bar{B}) = r_3 \quad p(C|\bar{A}, \bar{B}) = r_4;$$

with the probabilities of the complements given by normalization. With the additional two parameters to specify the Bernoulli probabilities $p(A) = r_5$ and $p(B) = r_6$, we have a total of 6 parameters.

Using the law of conditional probability and the expressions found in (a), we have that

$$\begin{aligned} r_1 &= p(C|A, B) = \frac{p(A, B, C)}{p(A, B)} = \frac{p_1}{p_1 + p_4}, \\ r_2 &= p(C|\bar{A}, B) = \frac{p(\bar{A}, B, C)}{p(\bar{A}, B)} = \frac{p_2}{p_2 + p_6}, \\ r_3 &= p(C|A, \bar{B}) = \frac{p(A, \bar{B}, C)}{p(A, \bar{B})} = \frac{p_3}{p_3 + p_5}. \end{aligned}$$

To compute r_4 , we use

$$r_4 = p(C|\bar{A}, \bar{B}) = \frac{p(\bar{A}, \bar{B}, C)}{p(\bar{A}, \bar{B})} = \frac{p(\bar{A}, \bar{B}, C)}{p(\bar{A}, \bar{B}, C) + p(\bar{A}, \bar{B}, \bar{C})} = \frac{p_7}{1 - p_1 - p_2 - p_3 - p_4 - p_5 - p_6}.$$

Finally, $r_5 = p(A) = q_5$ and $r_6 = p(B) = p(A, B) + p(\bar{A}, B) = p_1 + p_2 + p_4 + p_6$.

Solution to Exercise 3.2

The idea is to show that the joint density of A and B factorizes into two terms, using the joint density expressed by the Bayesian networks.

- a) We have that

$$p(A, B|C) = \frac{p(A, B, C)}{p(C)} = \frac{p(A|C)p(B|C)p(C)}{p(C)} = p(A|C)p(B|C),$$

so $A \perp\!\!\!\perp B|C$.

- b) We have that

$$\begin{aligned} p(A, B|C) &= \frac{p(A, B, C)}{p(C)} = \frac{1}{p(C)}p(B|C)p(C|A)p(A) \\ &= p(B|C)\frac{p(C|A)p(A)}{p(C)} = p(B|C)p(A|C), \end{aligned}$$

so $A \perp\!\!\!\perp B|C$.

- c) We have that

$$p(A, B) = \sum_C p(A, B, C) = \sum_C p(C|A, B)p(A)p(B) = p(A)p(B),$$

so $A \perp\!\!\!\perp B|\emptyset$.

Solution to Exercise 3.3

This exercise can be solved using the independence relationships expressed by the graph. Using d-separation, we have the following

- D is independent of the whole graph except F ;
- F is not independent of A, B , and E ;
- E is not independent of A, B , and F ;
- C is independent of the whole graph except E ;

Therefore, one possible Bayesian network resulting from the marginalization is presented in Figure 3.1.

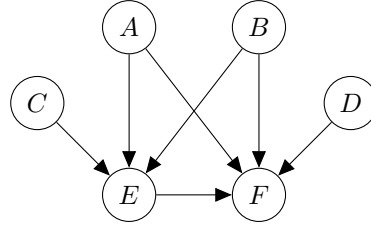


Figure 3.1: Bayesian network resulting from the marginalization of X in Figure 3.3. The edge between E and F has arbitrary orientation.

The same result can be achieved using explicit computations. To this end we let

$$p(A, B, C, D, E, F) = \sum_X p(A, B, C, D, E, F, X) = \sum_X p(F|D, X)p(E|C, X)p(X|A, B)p(A)p(B)p(C)p(D).$$

Now, using the dependency relations implied by the graph, we have that $E \perp\!\!\!\perp A, B|C, X$ and that $X \perp\!\!\!\perp C|A, B$, so

$$p(X|A, B) = p(X|A, B, C), \quad p(E|C, X) = p(E|A, B, C, X),$$

and

$$p(E|C, X)p(X|A, B) = p(E, X|A, B, C);$$

similarly, $F \perp\!\!\!\perp A, B, C, E|D, X$, so $p(F|D, X) = p(F|A, B, C, D, E, X)$ and

$$p(F|D, X)p(E, X|A, B, C) = p(F, E, X|A, B, C, D);$$

finally,

$$\begin{aligned} p(A, B, C, D, E, F) &= \sum_X p(F, E, X|A, B, C, D)p(A)p(B)p(C)p(D) \\ &= p(F, E|A, B, C, D)p(A)p(B)p(C)p(D). \end{aligned}$$

Using conditioning, we have that

$$p(F, E|A, B, C, D) = p(F|E, A, B, C, D)p(E|A, B, C, D);$$

note that, from the network, $F \perp\!\!\!\perp C|A, B, D$ and $E \perp\!\!\!\perp D|A, B, C$, so we have the result

$$p(A, B, C, D, E, F) = p(F|E, A, B, D)p(E|A, B, C)p(A)p(B)p(C)p(D).$$

Solution to Exercise 3.4

We compute the joint distribution $p(A, B)$ by using the law of total probability and marginalization:

$$p(A = 0, B = 0) = \sum_C p(A = 0, B = 0, C) = p(A = 0, B = 0, C = 0) + p(A = 0, B = 0, C = 1) = 0.336;$$

similarly, $p(A = 0, B = 1) = 0.264$, $p(A = 1, B = 0) = 0.256$, and $p(A = 1, B = 1) = 0.144$.

Using the same reasoning, we can compute the marginal distribution of A :

$$p(A = 0) = \sum_B p(A = 0, B) = p(A = 0, B = 0) + p(A = 0, B = 1) = 0.336 + 0.264 = 0.6;$$

and $p(A = 1) = 1 - p(A = 0) = 0.4$. And the marginal distribution of B :

$$p(B = 0) = \sum_A p(A, B = 0) = p(A = 0, B = 0) + p(A = 1, B = 0) = 0.336 + 0.256 = 0.592.$$

Finally $p(A = 0)p(B = 0) = 0.6 \cdot 0.592 = 0.3552 \neq p(A = 0, B = 0)$, So $A \not\perp\!\!\!\perp B$.

Consider now the marginal distribution of C , we have that

$$p(C = 0) = \sum_A \sum_B p(A, B, C = 0) = 0.48;$$

Hence,

$$p(A = 0, B = 0 | C = 0) = \frac{p(A = 0, B = 0, C = 0)}{p(C = 0)} = \frac{0.192}{0.48} = 0.4$$

hence,

$$p(A = 0 | C = 0) = \frac{p(A=0, C=0)}{p(C=0)} = \frac{p(A=0, B=0, C=0) + p(A=0, B=1, C=0)}{p(C=0)} = \frac{0.192 + 0.048}{0.48} = 0.5$$

and

$$p(B = 0 | C = 0) = \frac{p(B=0, C=0)}{p(C=0)} = \frac{p(A=0, B=0, C=0) + p(A=1, B=0, C=0)}{p(C=0)} = \frac{0.192 + 0.192}{0.48} = 0.8$$

Hence,

$$p(A = 0 | C = 0)p(B = 0 | C = 0) = 0.5 \cdot 0.8 = 0.4 = p(A = 0, B = 0 | C = 0).$$

Since similar relationships hold for the other possible outcomes, we have that $p(A, B | C) = p(A | C)p(B | C)$ and $A \perp\!\!\!\perp B | C$.

Solution to Exercise 3.5

Define the random variables Tuberculosis (T), Krakozhia (K), lung Damage (D), Lung cancer (L), Smoking (S), Bronchitis (B), and dysPnea (P). Figure 3.2 shows the final Bayesian network.

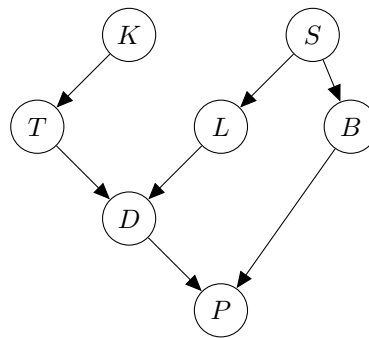


Figure 3.2: Bayesian network for the Krakozhia exercise. With the added chest X-ray variable.

- a) $T \perp\!\!\!\perp S | P$? No, they are connected by $T \rightarrow D \rightarrow P \rightarrow B \rightarrow S$.
- b) $L \perp\!\!\!\perp B | S$? Yes, as $L \rightarrow S \not\rightarrow B$ and $L \rightarrow D \rightarrow P \not\rightarrow B$
- c) $K \perp\!\!\!\perp S | L$? Yes, as $T \rightarrow D \not\rightarrow L$ and $T \rightarrow D \rightarrow P \not\rightarrow B$.
- d) $K \perp\!\!\!\perp S | L, P$? No, they are connected by $T \rightarrow D \rightarrow P \rightarrow B \rightarrow S$.

Solution to Exercise 3.6

Using (18)

- a) The Bayesian network is given in solid in Figure 3.3.
- b) $S \perp\!\!\!\perp A | C$? No, they are connected by $S \rightarrow C \rightarrow A$.
- c) See red addition in Figure 3.3.
- d) $W \perp\!\!\!\perp C | A$? Yes, because $W \rightarrow A \not\rightarrow C$

Solution to Exercise 3.7

Using (19)

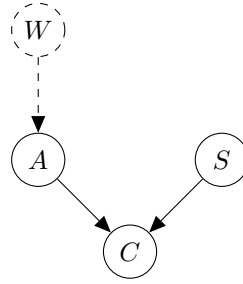


Figure 3.3: Bayesian network in the worker example.

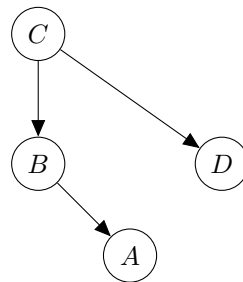


Figure 3.4: Bayesian network in 3.7

- a) The Bayesian network is given in Figure 3.4.
- b) The node C is tail-to-tail on the path from B to D , so we need to condition on this variable. We have that

$$\begin{aligned}
 p(B, D|C) &= \sum_A p(A, B, D|C) \\
 &= \frac{1}{p(C)} \sum_A p(A, B, D, C) \\
 &= \frac{1}{p(C)} \sum_A p(A|B)p(B|C)p(D|C)p(C) \\
 &= \sum_A p(A|B)p(B|C)p(D|C) \\
 &= p(B|C)p(D|C).
 \end{aligned}$$

Solution to Exercise 3.8

Define boolean random variables indicating whether there is light in the window (L), whether Marcus is in the office (O), and whether he is logged in to the mainframe (M). Laura’s model corresponds to the following probabilities:

$$p(O) = 0.8 \quad p(L|O) = 0.75 \quad p(L|\bar{O}) = 0.02 \quad p(M|O) = 0.92 \quad p(M|\bar{O}) = 0.15$$

- a) The Bayesian network of the model is given by Figure 3.5, and $p(M, O, L) = p(M|O)p(L|O)p(O)$.
- b) We have that

$$\begin{aligned}
 p(L) &= \sum_M \sum_O p(M, O, L) = \sum_M \sum_O p(M|O)p(L|O)p(O) \\
 &= \sum_O p(L|O)p(O) = 0.75 \cdot 0.8 + 0.02 \cdot 0.2 = 0.604;
 \end{aligned}$$

note that $\sum_M p(M|O) = 1$.

c) We have that

$$p(L|M) = \frac{p(L, M)}{p(M)},$$

where

$$\begin{aligned} p(L, M) &= \sum_O p(M, O, L) = p(M|O)p(L|O)p(O) + p(M|\bar{O})p(L|\bar{O})p(\bar{O}) \\ &= 0.92 \cdot 0.75 \cdot 0.8 + 0.15 \cdot 0.02 \cdot 0.2 = 0.5526 \end{aligned}$$

and

$$\begin{aligned} p(M) &= p(M|O)p(O) + p(M|\bar{O})p(\bar{O}) \\ &= 0.92 \cdot 0.8 + 0.15 \cdot 0.2 = 0.766; \end{aligned}$$

so $p(L|M) \approx 0.72 > p(L)$.

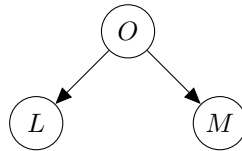


Figure 3.5: Bayesian network of the model in Exercise 3.8.

Solution to Exercise 3.9

Observing the conditional distribution in Table 3.2, we have that

$$p(E, O, I, S, F) = p(F|E, I)p(S|O)p(I|O, E)p(O|E)p(E)$$

a) The Bayesian network is presented in Figure 3.6.

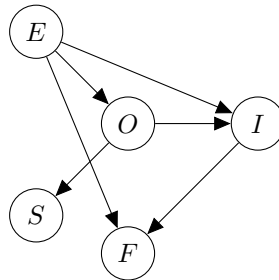


Figure 3.6: Bayesian network of the economic model in Exercise 3.9.

b) We are after $p(I = h|S = n, F = h)$; we can compute it using

$$p(I = h|S = n, F = h) = \frac{p(I = h, S = n, F = h)}{p(S = n, F = h)}.$$

We have that

$$\begin{aligned}
P(I = h, S = n, F = h) &= \sum_O \sum_E p(I = h, S = n, F = h, O, E) \\
&= \sum_O \sum_E p(F = h|O, E, I = h)p(S = n|O)p(I = h|O, E)p(O|E)p(E) \\
&= p(F = h|I = h, E = l)p(S = n|O = l)p(I = h|O = l, E = l)p(O = l|E = l)p(E = l) \\
&\quad + p(F = h|I = h, E = l)p(S = n|O = h)p(I = h|O = h, E = l)p(O = l|E = h)p(E = h) \\
&\quad + p(F = h|I = h, E = h)p(S = n|O = l)p(I = h|O = l, E = h)p(O = h|E = l)p(E = l) \\
&\quad + p(F = h|I = h, E = h)p(S = n|O = h)p(I = h|O = h, E = h)p(O = h|E = h)p(E = h) \\
&\approx 0.3096
\end{aligned}$$

and that

$$\begin{aligned}
P(S = n, F = h) &= \sum_I \sum_O \sum_E p(I, S = n, F = h, O, E) \\
&= p(I = h, S = n, F = h, O, E) + p(I = l, S = n, F = h, O, E) \\
&= p(I = h, S = n, F = h, O, E) + \sum_O \sum_E p(F = h|O, E, I = l)p(S = n|O)p(I = l|O, E)p(O|E)p(E) \\
&= p(I = h, S = n, F = h) \\
&\quad + p(F = h|I = l, E = l)p(S = n|O = h)p(I = l|O = h, E = l)p(O = l|E = h)p(E = h) \\
&\quad + p(F = h|I = l, E = h)p(S = n|O = l)p(I = l|O = l, E = h)p(O = h|E = l)p(E = l) \\
&\quad + p(F = h|I = l, E = h)p(S = n|O = h)p(I = l|O = h, E = h)p(O = h|E = h)p(E = h) \\
&\approx 0.3144
\end{aligned}$$

Hence,

$$p(I = h|S = n, F = h) = \frac{0.6902}{0.6985} \approx 0.9847$$

Bibliography

- [1] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- [2] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [3] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence*. CRC press, 2010.
- [4] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):157–194, 1988.
- [5] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.