School of Computing
National University of Singapore
CS5340: Uncertainty Modeling in AI
Semester 1, AY 2022/23

Exercise 1 - Solutions

Question 1

a)

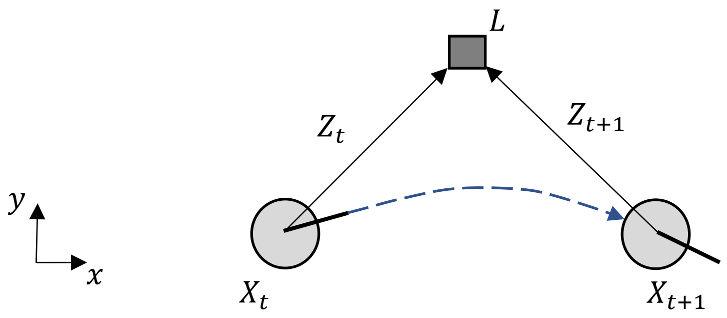


Fig. 1.1

Fig. 1.1 shows a mobile robot that traverses from pose X_t to X_{t+1} over time t to $t+1$. The robot is equipped with an 1-dimensional range sensor that returns the distances Z_t and Z_{t+1} of a landmark structure L in the environment from the poses X_t and X_{t+1} respectively. Let U_t denotes the control command given by the user to move the robot from X_t to X_{t+1}.

(i) Taking $\{U_t, L, X_t, X_{t+1}, Z_t, Z_{t+1}\}$ as random variables, state whether each of these random variables is an observed or latent/hidden random variable. Explain your answers.

Answer:

observed variables: $\{u_t, z_t, z_{t+1}\}$ (Inputs from user and observations from sensor);
latent variables: $\{x_t, x_{t+1}, l\}$

(ii) Given the following conditional independencies:

$$L \perp U_t \mid \emptyset, \quad X_t \perp L \mid U_t, \quad X_{t+1} \perp \{L, U_t\} \mid X_t,$$

$$Z_t \perp \{U_t, X_{t+1}\} \mid \{X_t, L\}, \quad Z_{t+1} \perp \{U_t, X_t, Z_t\} \mid \{L, X_{t+1}\}.$$

Write the factorized probability and draw the Bayesian network that represents the joint distribution $p(u_t, l, x_t, x_{t+1}, z_t, z_{t+1})$ assuming the following topological ordering of the random variables:

$$\{U_t, L, X_t, X_{t+1}, Z_t, Z_{t+1}\}.$$

Show all your workings clearly.

Answer:

From chain rule:

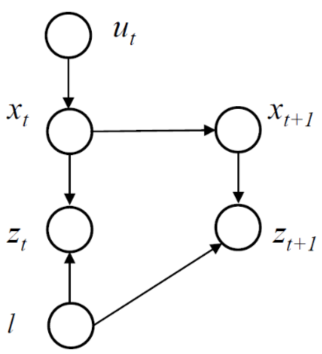
$$\begin{aligned} p(u_t, l, x_t, x_{t+1}, z_t, z_{t+1}) =& p(u_t)p(l|u_t)p(x_t|u_t, l)p(x_{t+1}|u_t, l, x_t) \\ & p(z_t|u_t, l, x_t, x_{t+1})p(z_{t+1}|u_t, l, x_t, x_{t+1}, z_t) \end{aligned}$$

$$\begin{aligned} p(l|u_t) = p(l) \quad & since \quad l \perp u_t|\emptyset \\ p(x_t|u_t, l) = p(x_t|u_t) \quad & since \quad x_t \perp l|u_t \\ p(x_{t+1}|u_t, l, x_t) = p(x_{t+1}|x_t) \quad & since \quad x_{t+1} \perp \{l, u_t\}|x_t \end{aligned}$$

$$\begin{aligned} p(z_t|u_t, l, x_t, x_{t+1}) = p(z_t|x_t, l) \quad & since \quad z_t \perp \{u_t, x_{t+1}\}|\{x_t, l\} \\ p(z_{t+1}|u_t, l, x_t, x_{t+1}, z_t) = p(z_{t+1}|l, x_{t+1}) \quad & since \quad z_{t+1} \perp \{u_t, x_t, z_t\}|\{l, x_{t+1}\} \end{aligned}$$

$$\Rightarrow p(u_t, l, x_t, x_{t+1}, z_t, z_{t+1}) = p(u_t)p(l)p(x_t|u_t)p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|l, x_{t+1})$$

Graphic model:



(iii) Write the following probability distribution $p(z_t, z_{t+1} \mid l)$ in terms of the factorized probability obtained in (ii). Simplify your answer.

Answer:

$$\boldsymbol{p(z_t, z_{t+1} \mid l) = \frac{p(z_t, z_{t+1}, l)}{p(l)}},$$

where

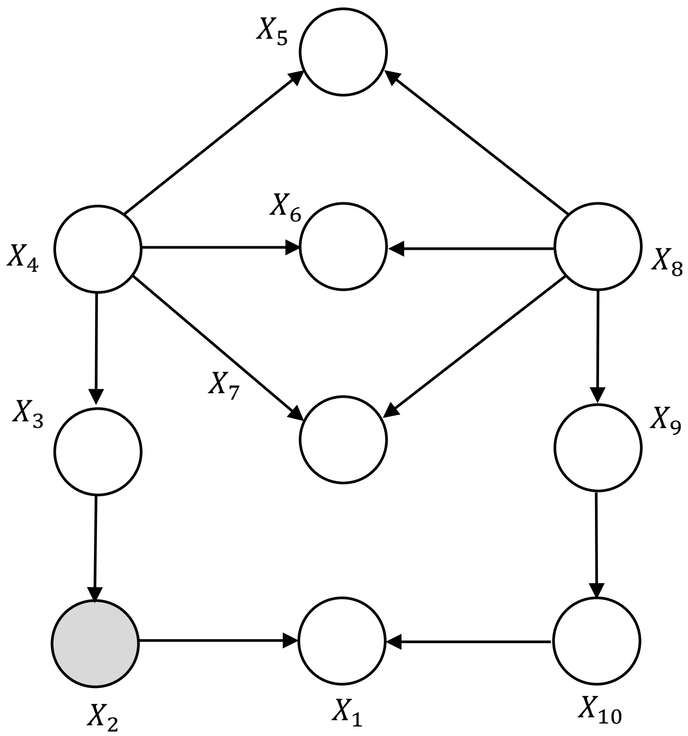
$$p(z_t, z_{t+1}, l) =$$
$$\sum_{u_t} \sum_{x_t} \sum_{x_{t+1}} p(u_t)p(l)p(x_t \mid u_t)p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l)$$
$$= p(l) \sum_{u_t} \sum_{x_t} \sum_{x_{t+1}} p(\cancel{u_t}) \frac{p(x_t, u_t)}{p(\cancel{u_t})} p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l)$$
$$= p(l) \sum_{x_t} \sum_{x_{t+1}} p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l) \sum_{u_t} p(x_t, u_t)$$
$$= p(l) \sum_{x_t} \sum_{x_{t+1}} p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l)p(x_t)$$

We get:

$$\boldsymbol{p(z_t, z_{t+1} \mid l) = \frac{p(z_t, z_{t+1}, l)}{p(l)}}$$
$$= \frac{\cancel{p(l)} \sum_{x_t} \sum_{x_{t+1}} p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l)p(x_t)}{\cancel{p(l)}}$$

$$= \sum_{x_t} \sum_{x_{t+1}} p(x_{t+1}|x_t)p(z_t|x_t, l)p(z_{t+1}|x_{t+1}, l)p(x_t)$$

b) For each of the Bayesian networks shown in Fig. 1.2, determine the largest set of nodes X_B such that $X_1 \perp X_B \mid X_2$. Explain your answers.



Answer:

(i) $x_1 \perp \{x_3, x_4\} \mid x_2$

Bayes ball can move to $x_{10}, x_9, x_8, x_7, x_6, x_5$, but blocked from $\{x_3, x_4\}$ because of head-to-head structures in $\{x_5, x_6, x_7\}$ (d-separation).

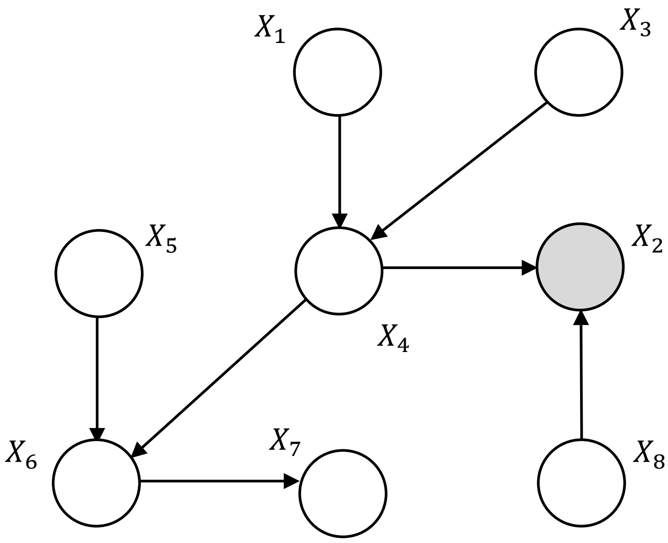


Fig. 1.2

Answer:

(ii) $x_1 \perp x_5 | x_2$

This is because of the head-to-head structure at $\{x_5 \rightarrow x_6 \leftarrow x_4\}$. Since x_6 is not observed, path is blocked from x_1 to x_5. Since x_2 is observed, path from x_1 to x_3 is not blocked because x_2 is a descendant of x_4. Similarly, the path from x_1 to x_8 is not blocked since x_2 is observed.

Question 2

Consider the graph shown in Fig. 2.1:

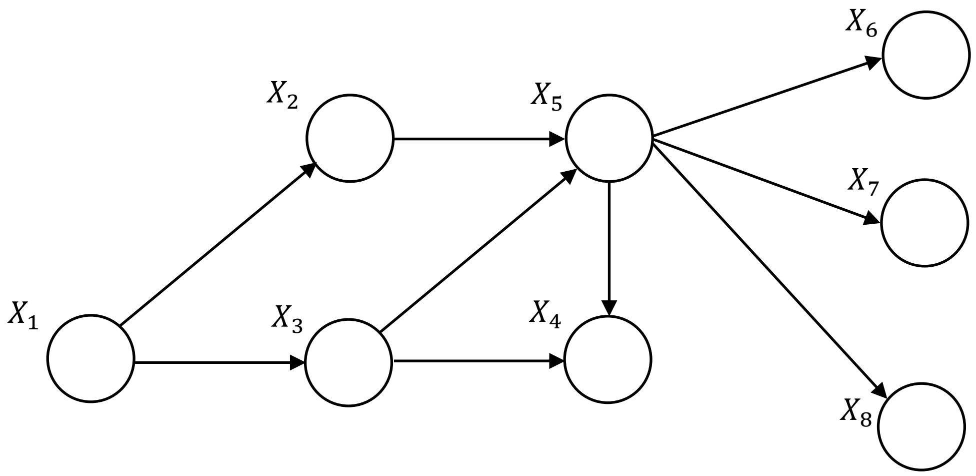
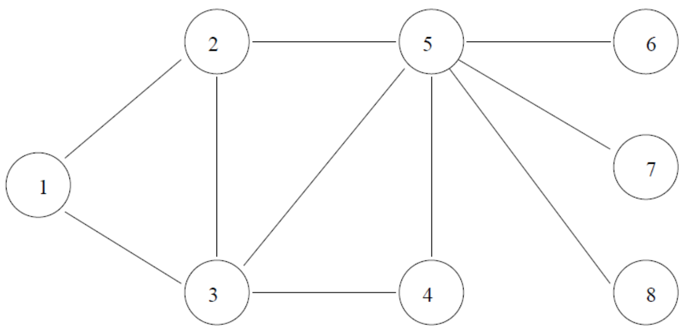


Fig 2.1

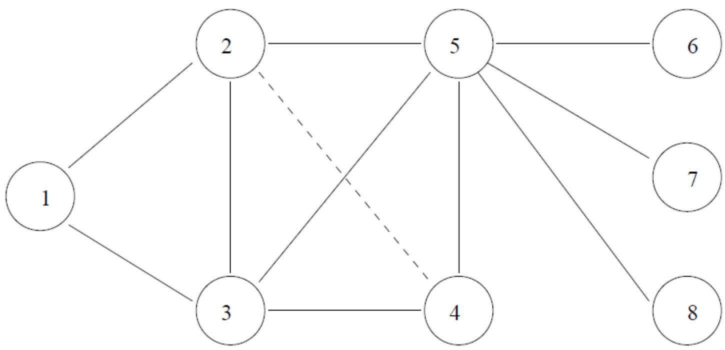
a) What is the corresponding moral graph?

Answer:



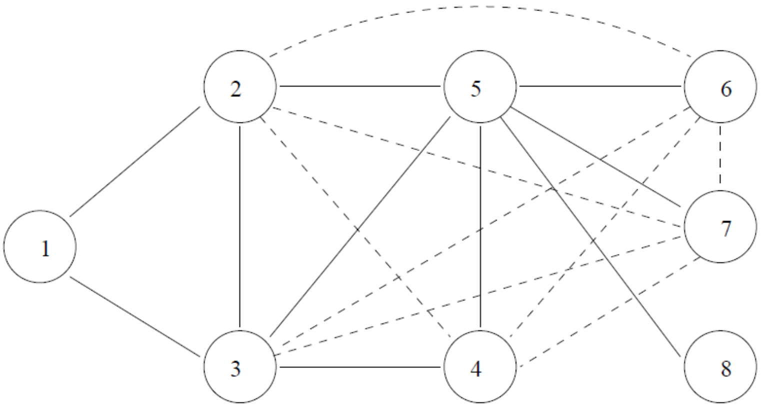
b) What is the reconstituted graph from the UNDIRECTEDGRAPHELIMINATE algorithm on the moral graph with the ordering {8,7,6,5,4,3,2,1}?

Answer:



c) What is the reconstituted graph from the UNDIRECTEDGRAPHELIMINATE algorithm on the moral graph with the ordering {8,5,6,7,4,3,2,1}?

Answer:



d) Suppose you wish to calculate $p(x_1|x_8)$. Which ordering is preferable? Why?

Answer:

The ordering in (b) is preferable because the maximum clique is of size 4 whereas the ordering in (c) results in a maximum clique of size 6.

Question 3

What is the treewidth of the graph below?

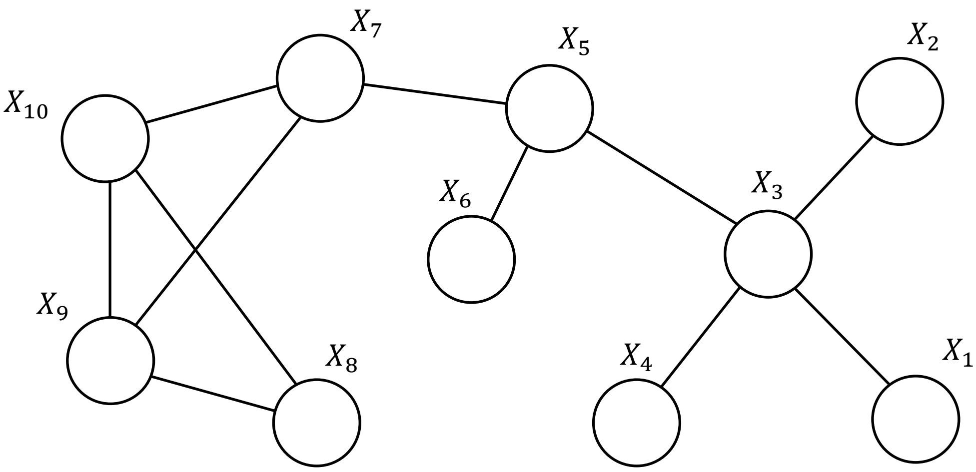


Fig 3.1

Answer:

- Elimination process adds new edges between (remaining) neighbors of the node, and this creates new "elimination cliques" in the graph.

- Treewidth: one less than the smallest achievable cardinality of the largest elimination clique over all possible elimination orderings.

We choose the elimination order that introduces the least "elimination cliques" into the graph, i.e. $I = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}\}$.

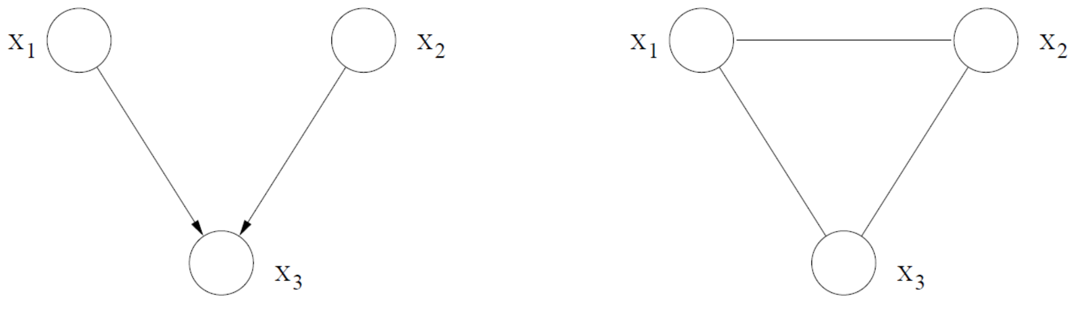
I introduces no new "elimination cliques" into the graph. We can see that the maximum clique is 3. Hence, the treewidth is 3-1 = 2.

Question 4

Consider the following random variables. X_1 and X_2 represent the outcomes of two independent fair coin tosses. X_3 is the indicator function of the event that the outcomes are identical.

a) Specify a directed graphical model that describes the joint probability distribution (i.e. specify the graph and the conditional distributions).

Answer:



(a) **(b)**

Figure 1: (a) The directed graphical model representing X_1, X_2, X_3 (b) An undirected graphical model representation for the same problem.

See figure (a) for the model. Let 1 denote heads and 0 denote tails. Then

$$P(X_1, X_2, X_3) = P(X_1)P(X_2)P(X_3|X_1, X_2).$$

where

$$P(X_1 = 1) = P(X_1 = 0) = \frac{1}{2},$$
$$P(X_2 = 1) = P(X_2 = 0) = \frac{1}{2},$$

$$P(X_3 = 0|X_1 = 0, X_2 = 0) = 0,$$
$$P(X_3 = 0|X_1 = 0, X_2 = 1) = 1,$$
$$P(X_3 = 0|X_1 = 1, X_2 = 0) = 1,$$
$$P(X_3 = 0|X_1 = 1, X_2 = 1) = 0,$$
$$P(X_3 = 1|X_1 = 0, X_2 = 0) = 1,$$
$$P(X_3 = 1|X_1 = 0, X_2 = 1) = 0,$$
$$P(X_3 = 1|X_1 = 1, X_2 = 0) = 0,$$
$$P(X_3 = 1|X_1 = 1, X_2 = 1) = 1,$$

b) Specify an undirected graphical model that describes the joint probability distribution (i.e. give the graph and specify the clique potentials).

Answer:

See figure (b) for the model. From part (a), there is a factor that includes all three

variables, hence there must be a corresponding clique with all three variables X_1, X_2 and X_3. Since there is only one clique, we have

$$P(X_1, X_2, X_3) = \frac{1}{Z}\Phi_1(X_1, X_2, X_3)$$

where $\Phi_1(X_1, X_2, X_3) = P(X_1, X_2, X_3) = P(X_1)P(X_2)P(X_3|X_1, X_2)$ and $Z = 1$.

c) In both cases, list all conditional independencies that are implied by the graph.

Answer:

From part (a), we can only state $X_1 \perp X_2|\varnothing$. For part (b), there is no conditional independence statements that is implied by the graph.

d) In both cases, list any additional conditional dependencies that are displayed by this probability distribution but are not implied by the graph.

Answer:

Other than $X_1 \perp X_2|\varnothing$, we also have $X_1 \perp X_3|\varnothing$ and $X_2 \perp X_3|\varnothing$. To see this, note that $P(X_1 = 0) = P(X_1 = 1) = P(X_2 = 0) = P(X_2 = 1) = P(X_3 = 0) = P(X_3 = 1) = 1/2$. We also have $P(x_1, x_2) = 1/4 = P(x_1)P(x_2)$ for all values of x_1, x_2. Similarly $P(x_2, x_3) = 1/4 = P(x_2)P(x_3)$ for all values of x_2, x_3.

This is a case where all three random variables are pairwise independent but are not mutually independent (i.e. $P(x_1, x_2, x_3) \neq P(x_1)P(x_2)P(x_3)$). Note also, that two of the conditional independence are numerical instead of structural and are easily destroyed by slight change in the parameter values i.e. if the coins are slightly unfair.

<u>Question 5</u>

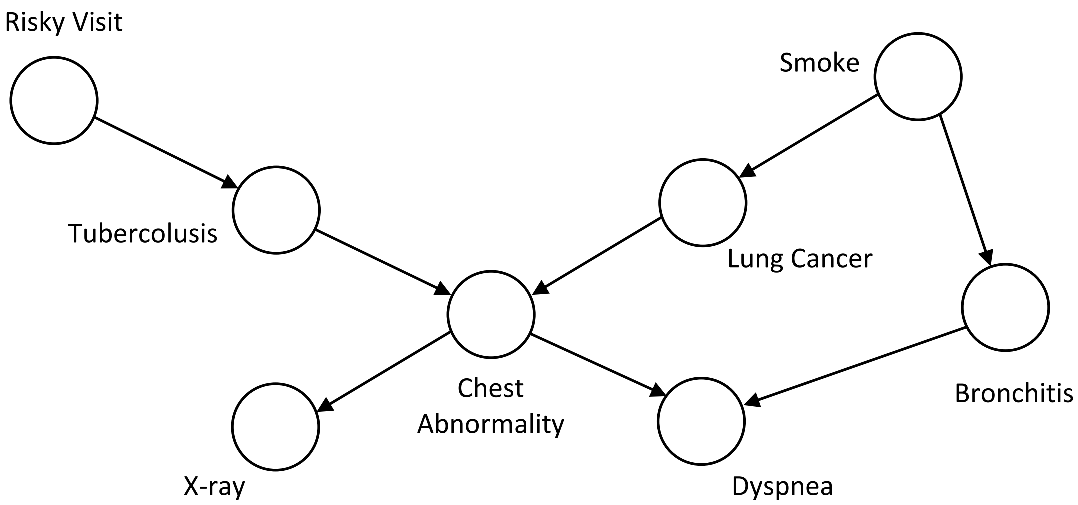


Fig. 5.1

The graphical model shown above describes some relationships among variables associated with chest abnormality. Answer the following questions based on the graphical model.

a) True or False. Justify your choice. $Smoke \perp Dyspnea \mid Bronchitis$.

   Answer:
   False. Although *Bronchitis* blocks one path from *Smoke* to *Dyspnea*, there is another path through *Lung Cancer* and *Chest Abnormality*.

b) True or False. Justify your choice. $Bronchitis \perp X\text{-}ray \mid Cancer$.

   Answer:
   True. One path is blocked by *Cancer* while another path is blocked by *Dyspnea*.

c) True or False. Justify your choice. $Smoke \perp Risky\,Visit \mid Dyspnea$.

Answer:
False. *Dyspnea* is a descendent of *Chest Abnormality* hence the path through *Tubercolusis*, *Chest Abnormality*, *Cancer* is no longer blocked by *Chest Abnormality*.

d) True of False. Justify your choice. *X-ray* $\perp$ $Smoke \mid \{Cancer, Bronchitis\}$.

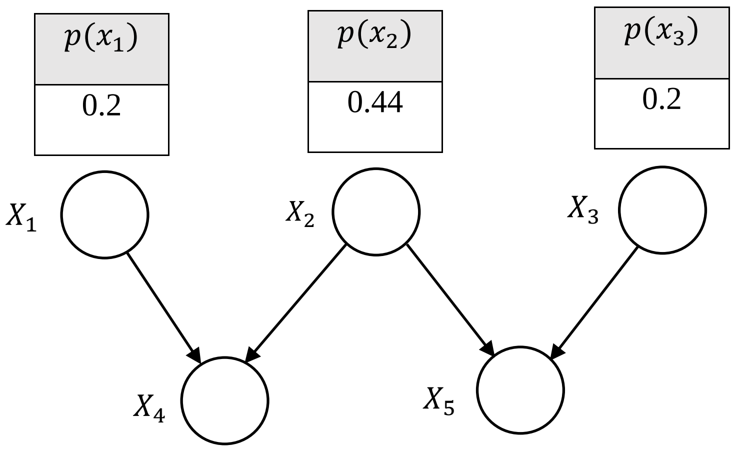
Answer:
True. One path is blocked by *Lung Cancer* while another is blocked by *Bronchitis*.

Question 6

Evaluate (give the distribution tables) the following probabilities:

$$p(x_1 \mid x_5), \quad p(x_2 \mid x_4), \quad p(x_3 \mid x_2), \quad p(x_4 \mid x_3), \quad p(x_5)$$

for the Bayesian network shown in Fig. 6.1, where each random variable takes a binary state, i.e. $x_i \in \{T, F\}$. Show all your workings clearly.



X_1	X_2	$p(x_4 \mid x_1, x_2)$
T	T	0.35
T	F	0.6
F	T	0.01
F	F	0.95

X_2	X_3	$p(x_5 \mid x_2, x_3)$
T	T	0.35
T	F	0.6
F	T	0.01
F	F	0.95

Fig. 6.1

Answer:

For $p(x_2 \mid x_4)$:

$$p(x_2|x_4) = \frac{p(x_2, x_4)}{\sum_{x_2} p(x_2, x_4)} = \frac{p(x_2, x_4)}{p(x_4)}$$

$$\begin{aligned} p(x_2, x_4) &= \sum_{x_1}\sum_{x_3}\sum_{x_5} p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2)p(x_5|x_2, x_3) \\ &= \sum_{x_1} p(x_1)p(x_2)p(x_4|x_1, x_2) \\ &= p(x_2)\sum_{x_1} p(x_1)p(x_4|x_1, x_2) \end{aligned}$$

$$p(x_2, x_4) = p(x_2)\{p(x_1 = 0)p(x_4|x_1 = 0, x_2) + p(x_1 = 1)p(x_4|x_1 = 1, x_2)\}$$

x_1	x_2	$p(x_4 = 1\|x_1, x_2)$	$p(x_4 = 0\|x_1, x_2)$
T	T	0.35	0.65
T	F	0.6	0.4
F	T	0.01	0.99
F	F	0.95	0.05

Given the following distribution tables:
We can compute:
1) $x_2 = 0, x_4 = 0$:

$$\begin{aligned} p(x_2, x_4) &= p(x_2 = 0)\{p(x_1 = 0)p(x_4 = 0|x_1 = 0, x_2 = 0) + p(x_1 = 1)p(x_4 = 0|x_1 = 1, x_2 = 0)\} \\ &= (0.56)\{(0.8)(0.05) + (0.2)(0.4)\} \\ &= 0.0672 \end{aligned}$$

2) $x_2 = 0, x_4 = 1$:

$$\begin{aligned} p(x_2, x_4) &= p(x_2 = 0)\{p(x_1 = 0)p(x_4 = 1|x_1 = 0, x_2 = 0) + p(x_1 = 1)p(x_4 = 1|x_1 = 1, x_2 = 0)\} \\ &= (0.56)\{(0.8)(0.95) + (0.2)(0.6)\} \\ &= 0.4928 \end{aligned}$$

3) $x_2 = 1, x_4 = 0$:

$$\begin{aligned} p(x_2, x_4) &= p(x_2 = 1)\{p(x_1 = 0)p(x_4 = 0|x_1 = 0, x_2 = 1) + p(x_1 = 1)p(x_4 = 0|x_1 = 1, x_2 = 1)\} \\ &= (0.44)\{(0.8)(0.99) + (0.2)(0.65)\} \\ &= 0.4057 \end{aligned}$$

4) $x_2 = 1, x_4 = 1$:

$$\begin{aligned} p(x_2, x_4) &= p(x_2 = 1)\{p(x_1 = 0)p(x_4 = 1|x_1 = 0, x_2 = 1) + p(x_1 = 1)p(x_4 = 1|x_1 = 1, x_2 = 1)\} \\ &= (0.44)\{(0.8)(0.01) + (0.2)(0.35)\} \\ &= 0.0343 \end{aligned}$$

Hence, the probability distribution of $p(x_2, x_4)$ is summarized as:

x_2	x_4	$p(x_2, x_4)$
0	0	0.0672
0	1	0.4928
1	0	0.4057
1	1	0.0343

Since

$$p(x_4) = \sum_{x_2} p(x_2) \sum_{x_1} p(x_1) p(x_4 | x_1, x_2),$$

$p(x_4) = \sum_{x_2} p(x_2, x_4) = p(x_2 = 0, x_4) + p(x_2 = 1, x_4)$

$p(x_4 = 0) = p(x_2 = 0, x_4 = 0) + p(x_2 = 1, x_4 = 0) = 0.0672 + 0.4057 = 0.4729$
$p(x_4 = 1) = p(x_2 = 0, x_4 = 1) + p(x_2 = 1, x_4 = 1) = 0.4928 + 0.0343 = 0.5271$

Since $\quad p(x_2|x_4) = \frac{p(x_2,x_4)}{p(x_4)}$,
if $x2 = 0, x_4 = 0, \quad p(x2|x4) = \frac{p(x_2=0,x_4=0)}{p(x_4=0)} = \frac{0.0672}{0.4729} = 0.1421;$

if $x2 = 0, x_4 = 1, \quad p(x2|x4) = \frac{p(x_2=0,x_4=1)}{p(x_4=1)} = \frac{0.4928}{0.5271} = 0.9349.$

For $p(x_1 | x_5), p(x_3 | x_2), \; p(x_4 | x_3)$:

$p(x_1|x_5) = p(x_1)$ since x_4 d-separates x_1 and x_5.
$p(x_3|x_2) = p(x_3)$ since x_5 d-separates x_2 and x_3.
$p(x_4|x_3) = p(x_4)$ since x_5 d-separates x_3 and x_4.

	$p(x_1)$	$p(x_2)$	$p(x_3)$
=1	0.2	0.44	0.2
=0	0.8	0.56	0.8

For $p(x_5)$:

$$p(x_5) = \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} p(x_1) p(x_2) p(x_3) p(x_4 | x_1, x_2) p(x_5 | x_2, x_3)$$
$$= \sum_{x_2} \sum_{x_3} p(x_2) p(x_3) p(x_5 | x_2, x_3)$$

x_2	x_3	$p(x_5 = 1 \vert x_2, x_3)$	$p(x_5 = 0 \vert x_2, x_3)$
T	T	0.35	0.65
T	F	0.6	0.4
F	T	0.01	0.99
F	F	0.95	0.05

x_2	x_3	$x_5 = 0$
0	0	$p(x_2 = 0)p(x_3 = 0)p(x_5 = 0 \mid x_2 = 0, x_3 = 0) =$ (0.56)(0.8)(0.05) = 0.0224
0	1	$p(x_2 = 0)p(x_3 = 1)p(x_5 = 0 \mid x_2 = 0, x_3 = 1) =$ (0.56)(0.2)(0.99) = 0.11088
1	0	$p(x_2 = 1)p(x_3 = 0)p(x_5 = 0 \mid x_2 = 1, x_3 = 0) =$ (0.44)(0.8)(0.40) = 0.1408
1	1	$p(x_2 = 1)p(x_3 = 1)p(x_5 = 0 \mid x_2 = 1, x_3 = 1) =$ (0.44)(0.2)(0.65) = 0.0572

x_2	x_3	$x_5 = 1$
0	0	$p(x_2 = 0)p(x_3 = 0)p(x_5 = 1 \mid x_2 = 0, x_3 = 0) =$ (0.56)(0.8)(0.95) = 0.4256
0	1	$p(x_2 = 0)p(x_3 = 1)p(x_5 = 1 \mid x_2 = 0, x_3 = 1) =$ (0.56)(0.2)(0.01) = 0.00112
1	0	$p(x_2 = 1)p(x_3 = 0)p(x_5 = 1 \mid x_2 = 1, x_3 = 0) =$ (0.44)(0.8)(0.60) = 0.2112
1	1	$p(x_2 = 1)p(x_3 = 1)p(x_5 = 1 \mid x_2 = 1, x_3 = 1) =$ (0.44)(0.2)(0.35) = 0.0308

x_5	$p(x_5)$
0	0.0224+0.11088+0.1408+0.0572=0.33128
1	0.4256+0.00112+0.2112+0.0308=0.66872

<u>Question 7</u>

Give the junction tree of the Bayesian network shown in Fig. 7.1 using the following elimination order: $\{X_7, X_6, X_5, X_4, X_3, X_2, X_1\}$. Show all your workings clearly.

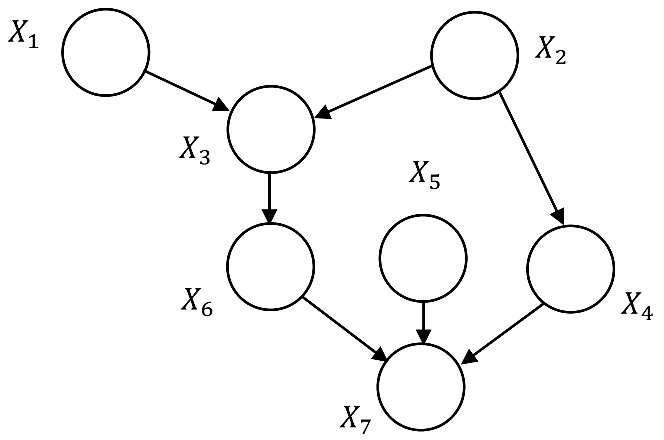
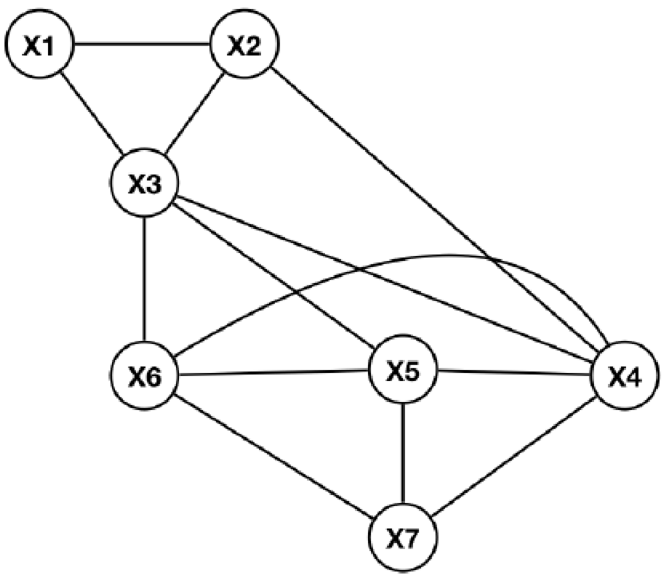


Fig. 7.1

Answer:

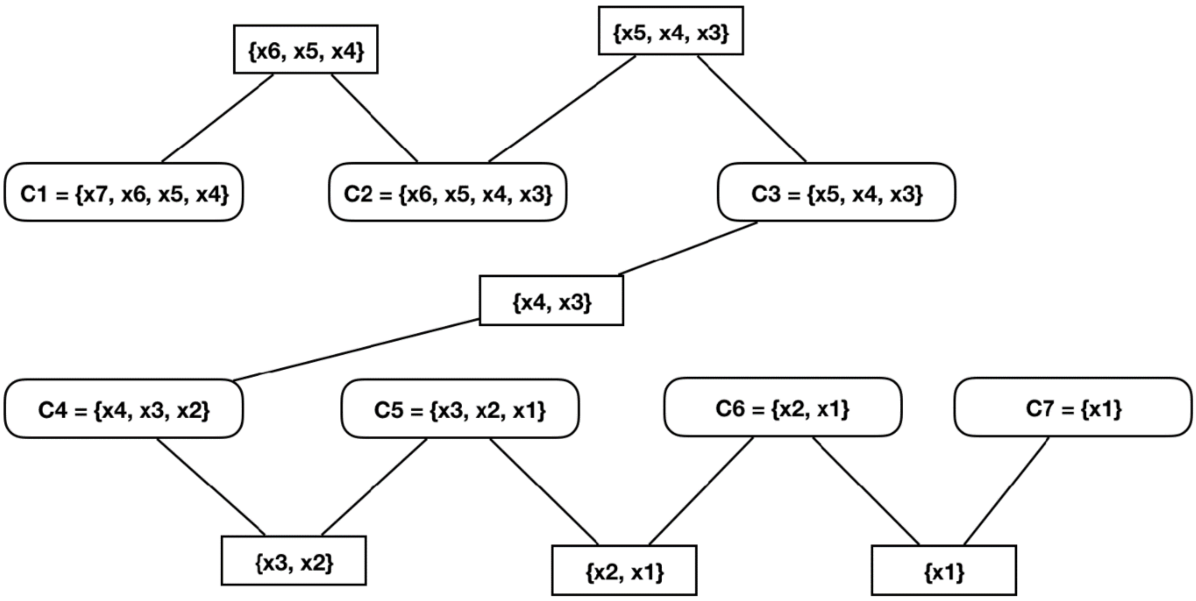
1. Moralization
2. Triangulation



3. Form clusters from elimination clusters:

$C_1 : \{x_7, x_6, x_5, x_4\}$
$C_2 : \{x_6, x_5, x_4, x_3\}$
$C_3 : \{x_5, x_4, x_3\}$
$C_4 : \{x_4, x_3, x_2\}$
$C_5 : \{x_3, x_2, x_1\}$
$C_6 : \{x_2, x_1\}$
$C_7 : \{x_1\}$

4. Get maximum spanning tree:



Question 8

Figure 8.1 shows a Bayesian network with five random variables X_1, X_2, X_3, X_4, X_5 , where $x_i \in \{0,1\}$ for $i = 1, 2, 4$, and $x_i \in \{0,1,2\}$ for $i = 3, 5$.

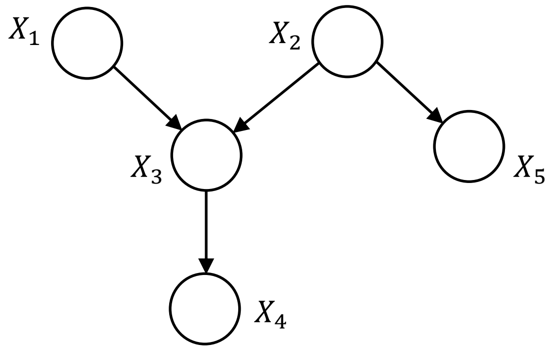


Figure 8.1

(a) Write down all the conditional independences given by the Bayesian network.

(b) Write down the factorized expression of the joint probability given by the Bayesian network.

(c) Convert the Bayesian network into a factor graph. Draw the factor graph and write down the expression of each factor clearly in your answer.

(d) Table 8.1 gives the probability tables of the Bayesian network, find the conditional probability $p(x_1|x_3 = 1, x_2)$. Show all your workings clearly.

X_1	X_2	X_3	$p(x_3\|x_1, x_2)$
0	0	0	0.3
0	0	1	0.4
0	1	0	0.9
0	1	1	0.08
1	0	0	0.05
1	0	1	0.25
1	1	0	0.5
1	1	1	0.3

X_1	$p(x_1)$
0	0.6

X_2	$p(x_2)$
0	0.7

X_3	X_4	$p(x_4\|x_3)$
0	0	0.1
1	0	0.4
2	0	0.99

Table 8.1

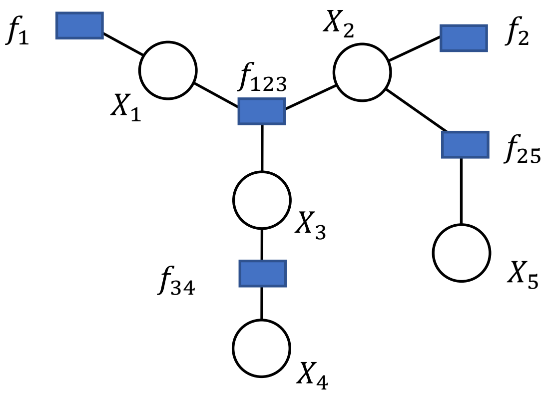
Answer:

a. $X_4 \perp \{X_1, X_2, X_5\} \mid X_3$
$X_5 \perp \{X_1, X_3, X_4\} \mid X_2$
$X_3 \perp X_5 \mid \{X_1, X_2\}$
$X_1 \perp X_2 \mid \emptyset$

$X_1 \perp \{X_2, X_5\} \mid \emptyset$

b. $p(X) = p(x_1)p(x_2)p(x_3|x_1,x_2)p(x_4|x_3)p(x_5|x_2)$

c.



Factors:

$f_1 = p(x_1)$

$f_2 = p(x_2)$

$f_{123} = p(x_3|x_1,x_2)$

$f_{34} = p(x_4|x_3)$

$f_{25} = p(x_5|x_2)$

d. $p(x_1|x_3 = 1, x_2) = \frac{p(x_1,x_2,x_3=1)}{p(x_2,x_3=1)}$,

$$\begin{aligned} p(x_1, x_2, x_3 = 1) &= \sum_{x_4} \sum_{x_5} p(x_1)p(x_2)p(x_3 = 1|x_1, x_2)p(x_4|x_3 = 1)p(x_5|x_2) \\ &= p(x_1)p(x_2)p(x_3 = 1|x_1, x_2)\underbrace{\boxed{\sum_{x_4} p(x_4|x_3 = 1) \sum_{x_5} p(x_5|x_2)}}_{= 1} \end{aligned}$$

X_1	X_2	$p(x_1, x_2, x_3 = 1) = p(x_1)p(x_2)p(x_3 = 1\|x_1, x_2)$
0	0	$p(x_1 = 0)p(x_2 = 0)p(x_3 = 1\|x_1 = 0, x_2 = 0) = (0.6)(0.7)(0.4) = \mathbf{0.168}$
0	1	$p(x_1 = 0)p(x_2 = 1)p(x_3 = 1\|x_1 = 0, x_2 = 1) = (0.6)(0.3)(0.08) = \mathbf{0.0144}$
1	0	$p(x_1 = 1)p(x_2 = 0)p(x_3 = 1\|x_1 = 1, x_2 = 0) = (0.4)(0.7)(0.25) = \mathbf{0.07}$
1	1	$p(x_1 = 1)p(x_2 = 1)p(x_3 = 1\|x_1 = 1, x_2 = 1) = (0.4)(0.3)(0.3) = \mathbf{0.036}$

$$\begin{aligned} p(x_2, x_3 = 1) &= \sum_{x_1} p(x_1, x_2, x_3 = 1) \\ &= p(x_2) \sum_{x_1} p(x_1)p(x_3 = 1|x_1, x_2) \end{aligned}$$

X_2	$p(x_2, x_3 = 1) = p(x_2) \sum_{x_1} p(x_1)p(x_3 = 1\|x_1, x_2)$
0	$p(x_2 = 0)\{p(x_1 = 0)p(x_3 = 1\|x_1 = 0, x_2 = 0) + p(x_1 = 1)p(x_3 = 1\|x_1 = 1, x_2 = 0)\}$ $= (0.7)\{(0.6)(0.4) + (0.4)(0.25)\} = \mathbf{0.238}$
1	$p(x_2 = 1)\{p(x_1 = 0)p(x_3 = 1\|x_1 = 0, x_2 = 1) + p(x_1 = 1)p(x_3 = 1\|x_1 = 1, x_2 = 1)\}$ $= (0.3)\{(0.6)(0.08) + (0.4)(0.3)\} = \mathbf{0.0504}$

$$p(x_1|x_3 = 1, x_2) = \frac{p(x_1, x_2, x_3 = 1)}{p(x_2, x_3 = 1)}$$

X_1	X_2	$p(x_1 \vert x_3 = 1, x_2) = \frac{p(x_1, x_2, x_3 = 1)}{p(x_2, x_3 = 1)}$
0	0	$\frac{p(x_1 = 0, x_2 = 0, x_3 = 1)}{p(x_2 = 0, x_3 = 1)} = \frac{0.168}{0.238} = \mathbf{0.706}$
0	1	$\frac{p(x_1 = 0, x_2 = 1, x_3 = 1)}{p(x_2 = 1, x_3 = 1)} = \frac{0.0144}{0.0504} = \mathbf{0.286}$
1	0	$\frac{p(x_1 = 1, x_2 = 0, x_3 = 1)}{p(x_2 = 0, x_3 = 1)} = \frac{0.07}{0.238} = \mathbf{0.294}$
1	1	$\frac{p(x_1 = 1, x_2 = 1, x_3 = 1)}{p(x_2 = 1, x_3 = 1)} = \frac{0.036}{0.0504} = \mathbf{0.714}$

Question 9

Figure 9.1 shows a graphical model with six binary-state latent random variables $Z = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}$, $z_i \in \{0,1\}$, and six binary-state observed random variables $X = \{X_1, X_2, X_3, X_4, X_5, X_6\}$, $x_i \in \{0,1\}$. Table 9.1 gives the pairwise potentials $\phi(z_i, z_j)$, $\forall\, ij \in \mathcal{E}_Z$ and conditional probability $p(x_i|z_i)$ for $i = 1, \dots 6$, where $\mathcal{E}_Z$ denotes all the edges between the latent random variables in the graphical model. Find the configuration of Z that maximizes the joint probability $p(X, Z)$.

(**Hint**: convert the graphical model into a factor graph, where the respective pairwise potential and conditional probability are represented as a single factor.)

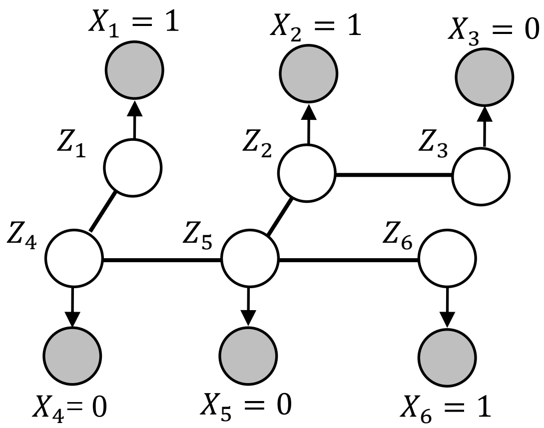


Figure 9.1

Z_i	Z_j	$\phi(z_i, z_j)$
0	0	0
0	1	2
1	0	2
1	1	0

X_i	Z_i	$p(x_i \mid z_i)$
0	0	0.9
0	1	0.05
1	0	0.1
1	1	0.95

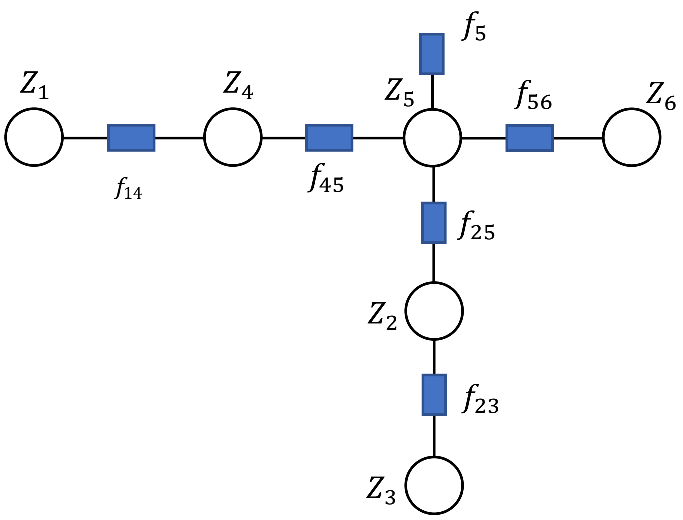
Table 9.1

Answer:

Joint probability:

$$p(X,Z) = \frac{1}{Z_p} p(x_1|z_1)p(x_2|z_2)p(x_3|z_3)p(x_4|z_4)p(x_5|z_5)p(x_6|z_6)\phi(z_1,z_4)\phi(z_4,z_5)\phi(z_2,z_5)\phi(z_2,z_3)\phi(z_5,z_6)$$

Convert the graphical model into a factor graph:



Factors:

$f_{14} = p(x_1 = 1|z_1)\phi(z_1, z_4)$

$f_{45} = p(x_4 = 0|z_4)\phi(z_4, z_5)$

$f_{23} = p(x_3 = 0|z_3)\phi(z_2, z_3)$

$f_{25} = p(x_2 = 1|z_2)\phi(z_2, z_5)$

$f_{56} = p(x_6 = 1|z_6)\phi(z_5, z_6)$

$f_5 = p(x_5 = 0|z_5)$

Apply max-product algorithm on factor graph:

$$m_{z_1}^{max}(z_4) = \max_{z_1} \mu_{f_{14}\to z_4} = \max_{z_1} p(x_1 = 1|z_1)\phi(z_1, z_4)$$

$\delta^{max}(z_1)$	Z_4	$m_{z_1}^{max}(z_4)$
1	0	1.9
0	1	0.2

Z_1	Z_4	$p(x_1 = 1|z_1)\phi(z_1, z_4)$
0	0	$0.1 \times 0 = 0$
0	1	$0.1 \times 2 = 0.2$
1	0	$0.95 \times 2 = 1.9$
1	1	$0.95 \times 0 = 0$

$m_{z_6}^{max}(z_5) = \max_{z_6} \mu_{f_{56} \to z_5} = \max_{z_6} p(x_6 = 1|z_6)\phi(z_5, z_6)$

Z_5	Z_6	$p(x_6 = 1\|z_6)\phi(z_5, z_6)$
0	0	$0.1 \times 0 = 0$
0	1	$0.95 \times 2 = 1.9$
1	0	$0.1 \times 2 = 0.2$
1	1	$0.95 \times 0 = 0$

$\delta^{max}(z_6)$	Z_5	$m_{z_6}^{max}(z_5)$
1	0	1.9
0	1	0.2

$m_{z_3}^{max}(z_2) = \max_{z_3} \mu_{f_{23} \to z_2} = \max_{z_3} p(x_3 = 0|z_3)\phi(z_2, z_3)$

Z_2	Z_3	$p(x_3 = 0\|z_3)\phi(z_2, z_3)$
0	0	$0.9 \times 0 = 0$
0	1	$0.05 \times 2 = 0.1$
1	0	$0.9 \times 2 = 1.8$
1	1	$0.05 \times 0 = 0$

$\delta^{max}(z_3)$	Z_2	$m_{z_3}^{max}(z_2)$
1	0	0.1
0	1	1.8

$m_{z_2}^{max}(z_5)$

$= \max_{z_2} \mu_{f_{25} \to z_5}$

$= \max_{z_2} p(x_2 = 1|z_2)\phi(z_2, z_5)m_{z_3}^{max}(z_2)$

Z_2	Z_5	$p(x_2 = 1\|z_2)\phi(z_2, z_5)m_{z_3}^{max}(z_2)$
0	0	$0.1 \times 0 \times 0.1 = 0$
0	1	$0.1 \times 2 \times 0.1 = 0.02$
1	0	$0.95 \times 2 \times 1.8 = 3.42$
1	1	$0.95 \times 0 \times 1.8 = 0$

$\delta^{max}(z_2)$	Z_5	$m_{z_2}^{max}(z_5)$
1	0	3.42
0	1	0.02

$m_{z_4}^{max}(z_5)$

$= \max_{z_4} \mu_{f_{45} \to z_5}$

$= \max_{z_4} p(x_4 = 0|z_4)\phi(z_4, z_5)\, m_{z_1}^{max}(z_4)$

Z_4	Z_5	$p(x_4 = 0\|z_4)\phi(z_4, z_5)m_{z_1}^{max}(z_4)$
0	0	$0.9 \times 0 \times 1.9 = 0$
0	1	$0.9 \times 2 \times 1.9 = 3.42$
1	0	$0.05 \times 2 \times 0.2 = 0.02$
1	1	$0.05 \times 0 \times 0.2 = 0$

$\delta^{max}(z_4)$	Z_5	$m_{z_4}^{max}(z_5)$
1	0	0.02
0	1	3.42

$m_{z_5}^{max}(z_5)$

$= \max_{z_5} m_{z_4}^{max}(z_5)\, m_{z_2}^{max}(z_5) m_{z_6}^{max}(z_5) p(x_5 = 0|z_5)$

Z_5	$m_{z_4}^{max}(z_5)m_{z_2}^{max}(z_5)m_{z_6}^{max}(z_5)p(x_5 = 0\|z_5)$
0	$0.02 \times 3.42 \times 1.9 \times 0.9 = 0.117$
1	$3.42 \times 0.02 \times 0.2 \times 0.05 = 0.000684$

Maximal probability = 0.117

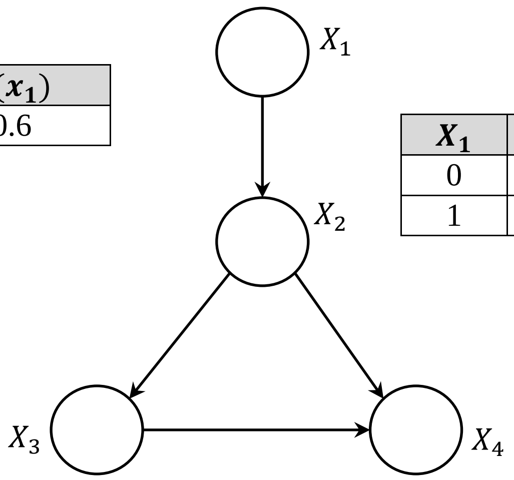
Backtrack to get all states that give the maximal probability:

$$\begin{tikzcd}
z_1^{max} = 0 & z_4^{max} = 1 \arrow[l] & z_5^{max} = 0 \arrow[l] \arrow[r] \arrow[d] & z_6^{max} = 1 \\
& & z_2^{max} = 1 \arrow[d] & \\
& & z_3^{max} = 0 &
\end{tikzcd}$$

Question 10

Figure 10.1 shows a Bayesian Network with four random variables X_1, X_2, X_3 and X_4, where $x_i \in \{0,1\}$. The respective prior and conditional probability distribution tables are also given.

$\boldsymbol{X_1}$	$\boldsymbol{p(x_1)}$
0	0.6



$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{p(x_2 \mid x_1)}$
0	0	0.6
1	0	0.3

$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{p(x_3 \mid x_2)}$
0	0	0.8
1	0	0.2

$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$	$\boldsymbol{p(x_4 \mid x_2, x_3)}$
0	0	0	0.1
0	1	0	0.8
1	0	0	0.5
1	1	0	0.6

Figure 10.1

Find the following marginal probabilities:

a. $p(x_2)$
b. $p(x_3)$

c. $p(x_4)$
d. $p(x_3, x_4)$

Answer:

Let $a = f(x_1, x_2) = p(x_1)p(x_2 \mid x_1)$ and $b = f(x_2, x_3, x_4) = p(x_3 \mid x_2)p(x_4 \mid x_2, x_3)$.

x1	p(x1)
0	0.6
1	0.4

x2	x3	p(x3|x2)
0	0	0.8
0	1	0.2
1	0	0.2
1	1	0.8

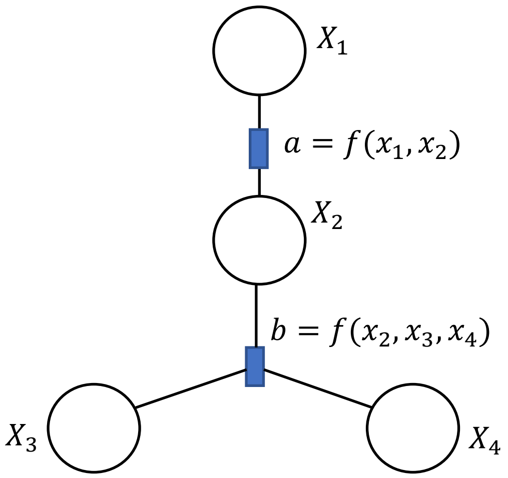
x1	x2	p(x2|x1)
0	0	0.6
0	1	0.4
1	0	0.3
1	1	0.7

x2	x3	x4	p(x4|x2,x3)
0	0	0	0.1
0	0	1	0.9
0	1	0	0.8
0	1	1	0.2
1	0	0	0.5
1	0	1	0.5
1	1	0	0.6
1	1	1	0.4

x1	x2	f(x1,x2)
0	0	0.36
0	1	0.24
1	0	0.12
1	1	0.28

x2	x3	x4	f(x2,x3,x4)
0	0	0	0.08
0	0	1	0.72
0	1	0	0.16
0	1	1	0.04
1	0	0	0.1
1	0	1	0.1
1	1	0	0.48
1	1	1	0.32

Factor graph:



Choosing X_2 as the root node:

Upward pass

$v_{1\to a} = 1, v_{3\to b} = 1, v_{4\to b} = 1,$

$u_{a\to 2}(x_2) = \sum_{x_1} f(x_1, x_2) = f(x_1 = 0, x_2) + f(x_1 = 1, x_2)$

$f(x_1 = 0, x_2 = 0) + f(x_1 = 1, x_2 = 0) = 0.36 + 0.12 = 0.48$

$f(x_1 = 0, x_2 = 1) + f(x_1 = 1, x_2 = 1) = 0.24 + 0.28 = 0.52$

$\boldsymbol{x_2}$	$\boldsymbol{u_{a\to 2}(x_2)}$
0	0.48
1	0.52

$u_{b\to 2}(x_2) = \sum_{x_3} \sum_{x_4} f(x_2, x_3, x_4)$

$= f(x_2, x_3 = 0, x_4 = 0) + f(x_2, x_3 = 0, x_4 = 1) +$

$f(x_2, x_3 = 1, x_4 = 0) + f(x_2, x_3 = 1, x_4 = 1)$

$f(x_2 = 0, x_3 = 0, x_4 = 0) + f(x_2 = 0, x_3 = 0, x_4 = 1) +$

$f(x_2 = 0, x_3 = 1, x_4 = 0) + f(x_2 = 0, x_3 = 1, x_4 = 1) = 0.08 + 0.72 + 0.16 + 0.04 = 1.0$

$f(x_2 = 1, x_3 = 0, x_4 = 0) + f(x_2 = 1, x_3 = 0, x_4 = 1) +$

$f(x_2 = 1, x_3 = 1, x_4 = 0) + f(x_2 = 1, x_3 = 1, x_4 = 1) = 0.1 + 0.1 + 0.48 + 0.32 = 1.0$

$\boldsymbol{x_2}$	$\boldsymbol{u_{b\to 2}(x_2)}$
0	1.0
1	1.0

a. $\boldsymbol{p(x_2) = u_{a\to 2}(x_2) \times u_{b\to 2}(x_2)}$

$\boldsymbol{x_2}$	$\boldsymbol{p(x_2)}$
0	0.48
1	0.52

Downward pass

$v_{2\to a} = u_{b\to 2}(x_2)$

$v_{2\to b} = u_{a\to 2}(x_2)$

$$u_{a\to 1}(x_1) = \sum_{x_2} v_{2\to a} \times a = \sum_{x_2} u_{b\to 2}(x_2) \times a = \sum_{x_2} 1.0 \times f(x_1, x_2)$$

$$= f(x_1, x_2 = 0) + f(x_1, x_2 = 1)$$

$$f(x_1 = 0, x_2 = 0) + f(x_1 = 0, x_2 = 1) = 0.36 + 0.24 = 0.6$$

$$f(x_1 = 1, x_2 = 0) + f(x_1 = 1, x_2 = 1) = 0.12 + 0.28 = 0.4$$

$\boldsymbol{x_1}$	$u_{a\to 1}(\boldsymbol{x_1})$
0	0.6
1	0.4

$$u_{b\to 3}(x_3) = \sum_{x_2}\sum_{x_4} v_{2\to b} \times b = \sum_{x_2}\sum_{x_4} u_{a\to 2}(x_2) \times b$$

$$= \sum_{x_2}\sum_{x_4} u_{a\to 2}(x_2) \times f(x_2, x_3, x_4)$$

$$= u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3, x_4 = 0) + u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3, x_4 = 1) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3, x_4 = 0) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3, x_4 = 1)$$

$$u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 0, x_4 = 0) + u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 0, x_4 = 1) +$$

$$u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 0, x_4 = 0) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 0, x_4 = 1) =$$

$$(0.48)(0.08) + (0.48)(0.72) + (0.52)(0.1) + (0.52)(0.1) = 0.488$$

$$u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 1, x_4 = 0) + u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 1, x_4 = 1) +$$

$$u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 1, x_4 = 0) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 1, x_4 = 1) =$$

$$(0.48)(0.16) + (0.48)(0.04) + (0.52)(0.48) + (0.52)(0.32) = 0.512$$

b. $\boldsymbol{p(x_3) = u_{b\to 3}(x_3)}$

$\boldsymbol{x_3}$	$\boldsymbol{u_{b\to 3}(x_3)}$
0	0.488
1	0.512

$$u_{b\to 4}(x_4) = \sum_{x_2}\sum_{x_3} v_{2\to b} \times b = \sum_{x_2}\sum_{x_3} u_{a\to 2}(x_2) \times b$$

$$= \sum_{x_2}\sum_{x_3} u_{a\to 2}(x_2) \times f(x_2, x_3, x_4)$$

$$= u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 0, x_4) + u_{a\to 2}(x_2 = 0) f(x_2 = 0, x_3 = 1, x_4) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 0, x_4) + u_{a\to 2}(x_2 = 1) f(x_2 = 1, x_3 = 1, x_4)$$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=0,x_4=0)+u_{a\to2}(x_2=0)f(x_2=0,x_3=1,x_4=0)+$

$u_{a\to2}(x_2=1)f(x_2=1,x_3=0,x_4=0)+u_{a\to2}(x_2=1)f(x_2=1,x_3=1,x_4=0)=$

$(0.48)(0.08)+(0.48)(0.16)+(0.52)(0.1)+(0.52)(0.48)=0.4168$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=0,x_4=1)+u_{a\to2}(x_2=0)f(x_2=0,x_3=1,x_4=1)+$

$u_{a\to2}(x_2=1)f(x_2=1,x_3=0,x_4=1)+u_{a\to2}(x_2=1)f(x_2=1,x_3=1,x_4=1)=$

$(0.48)(0.72)+(0.48)(0.04)+(0.52)(0.1)+(0.52)(0.32)=0.5832$

c. $\boldsymbol{p(x_4)=u_{b\to4}(x_4)}$

$\boldsymbol{x_4}$	$\boldsymbol{u_{b\to4}(x_4)}$
0	0.488
1	0.512

d. $p(x_3,x_4)=\sum_{x_2}v_{2\to b}\times f(x_2,x_3,x_4)=\sum_{x_2}u_{a\to2}(x_2)\times f(x_2,x_3,x_4)$

$$=u_{a\to2}(x_2=0)f(x_2=0,x_3,x_4)+u_{a\to2}(x_2=1)f(x_2=1,x_3,x_4)$$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=0,x_4=0)+u_{a\to2}(x_2=1)f(x_2=1,x_3=0,x_4=0)$

$=(0.48)(0.08)+(0.52)(0.1)=0.0904$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=0,x_4=1)+u_{a\to2}(x_2=1)f(x_2=1,x_3=0,x_4=1)$

$=(0.48)(0.72)+(0.52)(0.1)=0.3976$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=1,x_4=0)+u_{a\to2}(x_2=1)f(x_2=1,x_3=1,x_4=0)$

$=(0.48)(0.16)+(0.52)(0.48)=0.3264$

$u_{a\to2}(x_2=0)f(x_2=0,x_3=1,x_4=1)+u_{a\to2}(x_2=1)f(x_2=1,x_3=1,x_4=1)$

$=(0.48)(0.04)+(0.52)(0.32)=0.1856$

$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{p(x_3,x_4)}$
0	0	0.0904
0	1	0.3976
1	0	0.3264
1	1	0.1856

Question 11

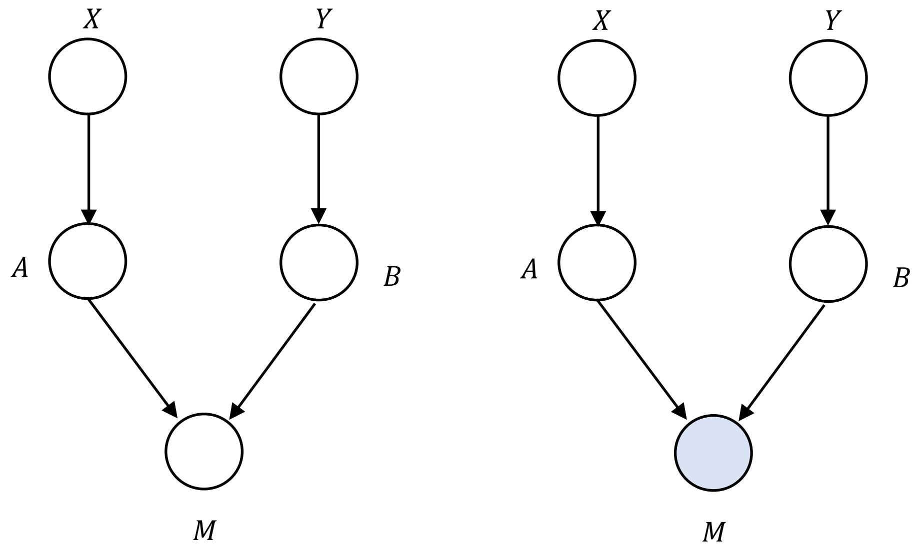
a) Assume that the day of the week that females are born on, X, is independent of the day of the week, Y, on which males are born. Assume, however, that the old rhyme is true, and that personality is dependent on the day of the week you're born on. If A represents the female personality type and B the male personality type, then $(A \top X)$ and $(B \top Y)$, but $(A \perp B)$. Whether or not a male and a female are married, M, depends strongly on their personality types, $(M \top A, B)$, but is independent of X and Y if we know A and B. Draw a graphical model that can represent this setting. What can we say about the (graphical) dependency between the days of the week that John and Jane are born on, given that they are not married?

Show all your workings clearly.

Note: $(A \top X)$ is the shorthand for A is dependent on X.

Answer:

$A \top X$; $B \top Y$; $A \perp B$ and $M \top \{A, B\}$.



In the case where M is observed, i.e. John and Jane are not married, we can see that A and B becomes dependent, i.e. we know that their personality types are not compatible. This implies that John and Jane are born in the day of the week that resulted in an incompatible personality type, respectively.

More technically, $A \rightarrow M \leftarrow B$ is a v-structure graph, hence, $A \not\perp B \mid M$. Since $A \top X$ and $B \top Y$, this implies $X \not\perp Y \mid M$.

b) Prove that the following relations are true or false:

i. $(X \perp Y)$ & $(Y \perp Z) \Longrightarrow (X \perp Z)$?

ii. $(X \perp Y \mid Z) \Longrightarrow (X \perp Y, W \mid Z)$?

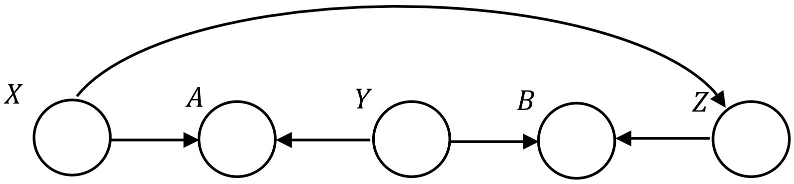
iii. Given that $(A \perp C \mid D, B)$ & $(A \perp B \mid D) \Longrightarrow (A \perp B, C \mid D)$, does $(A \perp B \mid D, C)$ & $(A \perp C \mid D, B) \Longrightarrow (A \perp B, C \mid D)$?

iv. $(X, Y, Z \perp A, B, C \mid D, E, F) \Longrightarrow (X \perp A, B \mid D, E, F)$ & $(X, Y \perp A \mid D, E, F)$ ..., i.e. $(any\ subset\ of\ \{X, Y, Z\} \perp any\ subset\ of\ \{A, B, C\} \mid D, E, F)$?

Show all your workings clearly.

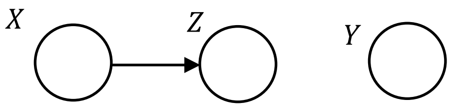
Answer:

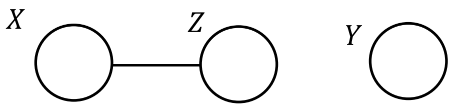
i. False. Counter-example:



$X \perp Y$ and $Y \perp Z$ but $X \not\perp Z$.

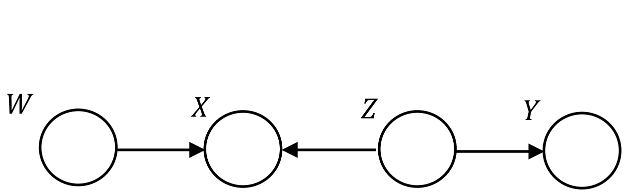
Other possible answers + many more…:





$\square$

ii. False. Counter-example:



$X \perp Y \mid Z$ but $X \not\perp YW \mid Z$

+ many more…

$\square$

iii. True. Given $(A \perp B \mid D, C)$ & $(A \perp C \mid D, B)$, we have $p(A \mid D, C) = p(A \mid B, D, C) = p(A \mid B, D)$. Apply product rule to first and last terms, we get

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

$$\Longrightarrow p(B \mid D) p(A, C \mid D) = p(C \mid D) p(A, B \mid D)$$

$$\Longrightarrow \sum_C p(B \mid D) p(A, C \mid D) = \sum_C p(C \mid D) p(A, B \mid D)$$

$\Longrightarrow p(B \mid D) p(A \mid D) = p(A, B \mid D)$, i.e. $(A \perp B \mid D)$

Hence $(A \perp B \mid D)$ and $(A \perp B \mid D, C) \Longrightarrow (A \perp B, C \mid D)$

□

iv. True. Let $\mathbb{M} = \{X, Y, Z\}$, $\mathbb{N} = \{A, B, C\}$ and $\mathbb{P} = \{D, E, F\}$, we can rewrite the conditional independence as $\mathbb{M} \perp \mathbb{N} \mid \mathbb{P}$, which gives us $p(\mathbb{M}, \mathbb{N} \mid \mathbb{P}) = p(\mathbb{M} \mid \mathbb{P}) p(\mathbb{N} \mid \mathbb{P})$. According to the d-separation or Markov property, this relation holds true for any subset of $\mathbb{M}$ and $\mathbb{N}$.

Alternatively,

$\{X, Y, Z \perp A, B, C\} \mid \{D, E, F\} \Longrightarrow p(X, Y, Z, A, B, C \mid DEF) = p(X, Y, Z \mid DEF) p(A, B, C \mid DEF)$, we can easily see that any subset of $\{X, Y, Z\}$ and $\{A, B, C\}$ can be achieved by marginalization, and the conditional independence still holds, e.g.

$$p(X, A, B \mid D, E, F) = \int_{Y,Z,C} p(X, Y, Z, A, B, C \mid D, E, F)$$

$$= \int_{Y,Z} p(X, Y, Z \mid D, E, F) \int_C p(A, B, C \mid D, E, F) = p(X \mid D, E, F) p(A, B \mid D, E, F)$$

$$\Longrightarrow X \perp \{A, B\} \mid \{D, E, F\}$$

□

Question 12

Figure 12.1 shows a directed graphical model with five random variables X_1, X_2, X_3, X_4, X_5, where $X_i \in \{0,1\}$. The respective conditional probabilities are shown in Table 12.1, where a, b, c and d are unknown values. Given that the minimal probability 0.00216 occurs at the configuration of $X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0$, and the maximal probability 0.10976 occurs at the configuration of $X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1$, find the **numerical values** of the probability distribution $p(X_2, X_3, X_5)$. **Show all your workings clearly**.

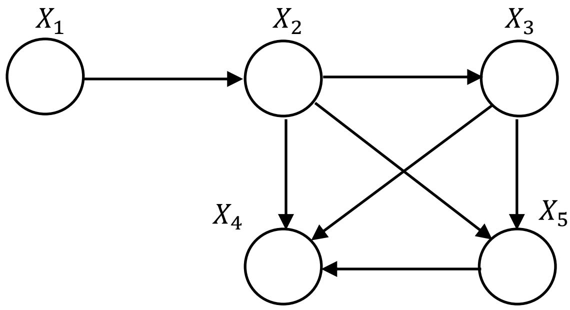


Figure 12.1.

X_2	X_3	X_5	$p(X_5\|X_2, X_3)$
0	0	0	0.3
0	0	1	0.7
0	1	0	0.1
0	1	1	0.9
1	0	0	0.5
1	0	1	0.5
1	1	0	0.8
1	1	1	0.2

X_2	X_3	$p(X_3\|X_2)$
0	0	0.7
0	1	0.3
1	0	0.6
1	1	0.4

X_1	X_2	$p(X_2\|X_1)$
0	0	0.8
0	1	0.2
1	0	c
1	1	d

X_4	X_2	X_3	X_5	$p(X_4\|X_2, X_3, X_5)$
0	0	0	0	0.2
0	0	0	1	0.3
0	0	1	0	0.6
0	0	1	1	0.9
0	1	0	0	0.2
0	1	0	1	0.3
0	1	1	0	0.2
0	1	1	1	0.6
1	0	0	0	0.8
1	0	0	1	0.7
1	0	1	0	0.4
1	0	1	1	0.1
1	1	0	0	0.8
1	1	0	1	0.7
1	1	1	0	0.8
1	1	1	1	0.4

X_1	$p(X_1)$
0	a
1	b

Table 12.1.

Answer:

Joint Probability:

$$p(x) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)p(x_4 \mid x_2, x_3, x_5)p(x_5 \mid x_2, x_3)$$

Given the minimal probability:

$p(x_1 = 1)p(x_2 = 0 \mid x_1 = 1)p(x_3 = 1 \mid x_2 = 0)p(x_4 = 1 \mid x_2 = 0, x_3 = 1, x_5 = 0)$
$p(x_5 = 0 \mid x_2 = 0, x_3 = 1) = 0.00216$, we have

$(b)(c)(0.3)(0.4)(0.1) = 0.00216 \Longrightarrow bc = 0.18$ ---(1)

Given the maximal probability:

$p(x_1 = 0)p(x_2 = 0 \mid x_1 = 0)p(x_3 = 0 \mid x_2 = 0)p(x_4 = 1 \mid x_2 = 0, x_3 = 0, x_5 = 1)$
$p(x_5 = 1 \mid x_2 = 0, x_3 = 0) = 0.10976$, we have

$(a)(0.8)(0.7)(0.7)(0.7) = 0.10976 \Longrightarrow a = 0.4$ ---(2)

Using a + b = 1 to solve for b in eq (2), we get $b = 1 - 0.4 = 0.6$.

Putting b into eq (1), we get $c = \frac{0.18}{0.6} = 0.3$.

Solving for

$p(X_2, X_3, X_5)$

$= \sum_{X_1} \sum_{X_4} p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2)p(x_4 \mid x_2, x_3, x_5)p(x_5 \mid x_2, x_3)$

$= p(x_3 \mid x_2)p(x_5 \mid x_2, x_3) \sum_{X_1} p(x_1)p(x_2 \mid x_1)$

$= p(x_3 \mid x_2)p(x_5 \mid x_2, x_3)[p(x_1 = 0)p(x_2 \mid x_1 = 0) + p(x_1 = 1)p(x_2 \mid x_1 = 1)]$

X_2	X_3	X_5	$p(X_2, X_3, X_5)$
0	0	0	$p(x_3 = 0 \mid x_2 = 0)p(x_5 = 0 \mid x_2 = 0, x_3 = 0)[p(x_1 = 0)p(x_2 = 0 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 0 \mid x_1 = 1)]$ $= (0.7)(0.3)[(0.4)(0.8) + (0.6)(0.3)] = \mathbf{0.105}$
0	0	1	$p(x_3 = 0 \mid x_2 = 0)p(x_5 = 1 \mid x_2 = 0, x_3 = 0)[p(x_1 = 0)p(x_2 = 0 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 0 \mid x_1 = 1)]$ $= (0.7)(0.7)[(0.4)(0.8) + (0.6)(0.3)] = \mathbf{0.245}$
0	1	0	$p(x_3 = 1 \mid x_2 = 0)p(x_5 = 0 \mid x_2 = 0, x_3 = 1)[p(x_1 = 0)p(x_2 = 0 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 0 \mid x_1 = 1)]$ $= (0.3)(0.1)[(0.4)(0.8) + (0.6)(0.3)] = \mathbf{0.015}$
0	1	1	$p(x_3 = 1 \mid x_2 = 0)p(x_5 = 1 \mid x_2 = 0, x_3 = 1)[p(x_1 = 0)p(x_2 = 0 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 0 \mid x_1 = 1)]$ $= (0.3)(0.9)[(0.4)(0.8) + (0.6)(0.3)] = \mathbf{0.135}$
1	0	0	$p(x_3 = 0 \mid x_2 = 1)p(x_5 = 0 \mid x_2 = 1, x_3 = 0)[p(x_1 = 0)p(x_2 = 1 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 1 \mid x_1 = 1)]$ $= (0.6)(0.5)[(0.4)(0.2) + (0.6)(0.7)] = \mathbf{0.15}$
1	0	1	$p(x_3 = 0 \mid x_2 = 1)p(x_5 = 1 \mid x_2 = 1, x_3 = 0)[p(x_1 = 0)p(x_2 = 1 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 1 \mid x_1 = 1)]$ $= (0.6)(0.5)[(0.4)(0.2) + (0.6)(0.7)] = \mathbf{0.15}$
1	1	0	$p(x_3 = 1 \mid x_2 = 1)p(x_5 = 0 \mid x_2 = 1, x_3 = 1)[p(x_1 = 0)p(x_2 = 1 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 1 \mid x_1 = 1)]$ $= (0.4)(0.8)[(0.4)(0.2) + (0.6)(0.7)] = \mathbf{0.16}$
1	1	1	$p(x_3 = 1 \mid x_2 = 1)p(x_5 = 1 \mid x_2 = 1, x_3 = 1)[p(x_1 = 0)p(x_2 = 1 \mid x_1 = 0)$ $+ p(x_1 = 1)p(x_2 = 1 \mid x_1 = 1)]$ $= (0.4)(0.2)[(0.4)(0.2) + (0.6)(0.7)] = \mathbf{0.04}$

Question 13

Prove the following conditional independences are true:

a) $(X \perp Y, W \mid Z) \Longrightarrow (X \perp Y \mid Z)$. This is also known as **Decomposition**.

b) $(X \perp Y, W \mid Z) \Longrightarrow (X \perp Y \mid Z, W)$. This is also known as **Weak Union**.

c) $(X \perp Y \mid Z)$ and $(X \perp W \mid Z, Y) \Longrightarrow (X \perp W, Y \mid Z)$. This is also known as **Contraction**.

Answer:

a)

$$p(X, Y \mid Z) = \sum_W p(X, Y, W \mid Z)$$

$$= \sum_W p(X \mid Z)p(Y, W \mid Z) \quad \text{since } (X \perp Y, W \mid Z)$$

$$= p(X \mid Z)\sum_W p(Y, W \mid Z)$$

$$= p(X \mid Z)p(Y \mid Z) \Longrightarrow (X \perp Y \mid Z)$$

i.e. $(X \perp Y, W \mid Z) \Longrightarrow (X \perp Y \mid Z)$.

b)

$$p(X,Y,W \mid Z) = p(X \mid Z)p(Y,W \mid Z) \quad \text{since } (X \perp Y,W \mid Z)$$

$$= p(X \mid Z)p(Y \mid Z,W)p(W \mid Z) \quad \text{(chain rule)}$$

$$= p(X \mid Z,W)p(Y \mid Z,W)p(W \mid Z)$$

We can write $p(X \mid Z) = p(X \mid Z,W)$ since $p(X \mid Z)p(W \mid Z)$ implies $(X \perp W \mid Z)$. The factorization of $p(X \mid Z,W)p(Y \mid Z,W)$ implies $(X \perp Y \mid Z,W)$.

Therefore, $(X \perp Y,W \mid Z) \Longrightarrow (X \perp Y \mid Z,W)$.

c)

$$p(X,Y,W \mid Z) = p(X \mid W,Y,Z)p(W,Y \mid Z) \quad \text{(chain rule)}$$

$$= p(X \mid Y,Z)p(W,Y \mid Z) \quad \text{since } (X \perp W \mid Z,Y)$$

$$= p(X \mid Z)p(W,Y \mid Z) \quad \text{since } (X \perp Y \mid Z)$$

i.e. $(X \perp Y \mid Z)$ and $(X \perp W \mid Z,Y) \Longrightarrow (X \perp W,Y \mid Z)$.

Question 14

Figure 4.1 shows a three-node Bayesian network with random variables $X_i \in \mathbb{R}, i = 1,2,3$. Furthermore, $p(X_1 \mid \mu_1, \sigma_1^2) = \mathcal{N}(X_1 \mid \mu_1, \sigma_1^2)$, $p(X_2 \mid \mu_2, \sigma_2^2) = \mathcal{N}(X_2 \mid \mu_2, \sigma_2^2)$, and $p(X_3 \mid X_1, X_2, w_0, w_1, w_2, \sigma_3^2) = \mathcal{N}(X_3 \mid w_0 + w_1X_1 + w_2X_2, \sigma_3^2)$, where $\mathcal{N}(X \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5 \frac{(X-\mu)^2}{\sigma^2}\right\}$ is the Gaussian distribution parameterized by the mean μ and variance σ^2. We have the linear-Gaussian distribution when the mean μ is parameterized by $w_0, \dots w_M$ as a weighted sum of the parent nodes $X_{\pi,1} \dots X_{\pi,M}$ of X, i.e., $\mu = w_0 + \sum_m w_m X_{\pi,m}$. Given 10 observations of the three-node Bayesian network in Table 4.1, find the probability of $p(X_1 = 1.5, X_2 = 0.5, X_3 = 1.0)$. **Show all your workings clearly.**

Useful equation: $\frac{d}{dx}\ln\big(f(x)\big) = \frac{1}{f(x)}\frac{df(x)}{dx}$.

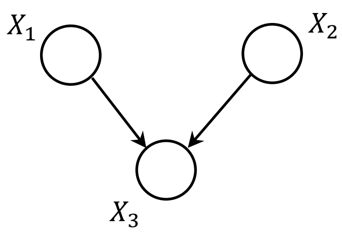


Figure. 4.1

Observation #	X_1	X_2	X_3
1	1.86	-0.03	-0.31
2	1.68	1.09	1.48
3	2.41	-0.45	2.18
4	2.40	-0.33	2.98
5	0.87	0.99	3.08
6	1.96	2.83	2.24
7	1.78	0.07	-0.89
8	2.81	1.44	1.18
9	3.42	0.72	0.88
10	3.44	2.34	6.15

Table. 4.1

(*25 marks*)

Solution

Joint probability:

$$p(X_1, X_2, X_3) = \ p(X_1 \mid \mu_1, \sigma_1^2)p(X_2 \mid \mu_2, \sigma_2^2)p(X_3 \mid X_1, X_2, w_0, w_1, w_2, \sigma_3^2)$$

$$\ln p(X_1, X_2, X_3) = \ln p(X_1 \mid \mu_1, \sigma_1^2) + \ln p(X_2 \mid \mu_2, \sigma_2^2) + \ln p(X_3 \mid X_1, X_2, w_0, w_1, w_2, \sigma_3^2)$$

$$\text{argmax}_{\mu_1, \sigma_1^2} \ln p(X_1, X_2, X_3) = \ \text{argmax}_{\mu_1, \sigma_1^2} \sum_n \ln p\left(X_{1,n} \mid \mu_1, \sigma_1^2 \right)$$

$$= \text{argmax}_{\mu_1, \sigma^2{}_1} \ \sum_n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-0.5 \ \frac{\left(X_{1,n} - \mu_1\right)^2}{\sigma_1^2}\right\}$$

$$= \text{argmax}_{\mu_1, \sigma_1^2} \sum_n \left\{- \frac{\left(X_{1,n} - \mu_1\right)^2}{\sigma_1^2} - 0.5 \ln 2\pi - 0.5 \ln \sigma^2\right\}.$$

$$\mathcal{L} = \sum_n \left\{-0.5 \frac{\left(X_{1,n} - \mu_1\right)^2}{\sigma_1^2} - 0.5 \ln 2\pi - 0.5 \ln \sigma^2\right\}$$

$$\frac{\delta\mathcal{L}}{\delta\mu_1} = \frac{\sum_n (X_{1,n} - \mu_1)}{\sigma_1^2} = 0 \ \Rightarrow \ \sum_n \left(X_{1,n} - \mu_1\right) = 0 \ \Rightarrow \ \boldsymbol{\mu_1} = \frac{\sum_n \boldsymbol{X_{1,n}}}{\boldsymbol{N_1}}$$

$$\frac{\delta\mathcal{L}}{\delta\sigma_1^2} = \sum_n \left\{0.5 \frac{\left(X_{1,n} - \mu_1\right)^2}{\sigma_1^4} - 0.5 \frac{1}{\sigma_1^2}\right\} = 0 \ \Rightarrow \boldsymbol{\sigma_1^2} = \frac{\sum_n \left(\boldsymbol{X_{1,n}} - \boldsymbol{\mu_1}\right)^2}{\boldsymbol{N_1}} \quad .$$

$$\boldsymbol{\mu_1} = \frac{\sum_n \boldsymbol{X_{1,n}}}{\boldsymbol{N_1}} = \frac{\mathbf{1.86 \ + \ 1.68 \ + \ 2.41 \ + \ 2.40 \ + \ 0.87 \ + \ 1.96 \ + \ 1.78 \ + \ 2.81 \ + \ 3.42 \ + \ 3.44}}{\mathbf{10}}$$

$$= \mathbf{22.63}$$

$$\boldsymbol{\sigma_1^2 = \frac{\sum_n \left(X_{1,n} - \mu_1\right)^2}{N_1}}$$

$$= \frac{(1.86-2.263)^2+(1.68-2.263)^2+(2.41-2.263)^2+(2.40-2.263)^2+(0.87-2.263)^2+(1.96-2.263)^2+(1.78-2.263)^2+(2.81-2.263)^2+(3.42-2.263)^2+(3.44-2.263)^2}{10}$$

$$= \mathbf{0.5831}$$

$$\boldsymbol{\mu_2 = \frac{\sum_n X_{2,n}}{N_2} = \ \ 0.8670,}$$

$$\boldsymbol{\sigma_2^2 = \frac{\sum_n \left(X_{2,n} - \mu_2\right)^2}{N_2} = 1.1045.}$$

$$\operatorname{argmax}_{w_0, w_1, w_2, \sigma_3^2} \ln p(X_1, X_2, X_3) =$$
$$\operatorname{argmax}_{w_0, w_1, w_2, \sigma_3^2} \textstyle\sum_n \ln p\big(\, X_{3,n} \mid X_{1,n}, X_{2,n}, w_0, w_1, w_2, \sigma_3^2 \,\big)$$

$$\mathcal{L} = \textstyle\sum_n \left\{ -0.5 \ \frac{\left(X_{3,n} - w_0 - w_1 X_{1,n} - w_2 X_{2,n}\right)^2}{\sigma_3^2} - 0.5 \ln 2\pi - 0.5 \ln \sigma_3^2 \right\}$$

$$\frac{\delta \mathcal{L}}{\delta w_0} = \textstyle\sum_n X_{3,n} - w_0 - w_1 X_{1,n} - w_2 X_{2,n} \ = 0$$

$$\frac{\delta \mathcal{L}}{\delta w_1} = \textstyle\sum_n (X_{3,n} - w_0 - w_1 X_{1,n} - w_2 X_{2,n}) X_{1,n} \ = 0$$

$$\frac{\delta \mathcal{L}}{\delta w_2} = \textstyle\sum_n (X_{3,n} - w_0 - w_1 X_{1,n} - w_2 X_{2,n}) X_{2,n} \ = 0$$

$$\begin{bmatrix} N & \sum_n X_{1,n} & \sum_n X_{2,n} \\ \sum_n X_{1,n} & \sum_n X_{1,n} X_{1,n} & \sum_n X_{2,n} X_{1,n} \\ \sum_n X_{2,n} & \sum_n X_{1,n} X_{2,n} & \sum_n X_{2,n} X_{2,n} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \sum_n X_{3,n} \\ \sum_n X_{3,n} X_{1,n} \\ \sum_n X_{3,n} X_{2,n} \end{bmatrix}$$

$$\begin{bmatrix} 10 & 22.63 & 8.67 \\ 22.63 & 57.0431 & 20.99 \\ 8.67 & 20.99 & 18.5619 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 18.97 \\ 47.2828 \\ 25.708 \end{bmatrix}$$

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -0.0502 \\ 0.5661 \\ 0.7683 \end{bmatrix}$$

$$\frac{\delta \mathcal{L}}{\delta \sigma_3^2} = \sum_n \left\{ 0.5 \ \frac{\left(X_{3,n} - w_0 - w_1 X_{1,n} - w_2 X_{2,n}\right)^2}{\sigma_3^4} - 0.5 \frac{1}{\sigma_3^2} \right\} = 0$$

$$\boldsymbol{\sigma_3^2} = \frac{\sum_{\boldsymbol{n}} \left(\boldsymbol{X_{3,n}} - w_0 - w_1 X_{1,n} - w_2 X_{2,n}\right)^2}{\boldsymbol{N}} = \mathbf{2.5639}$$

$p(X_1 = 1.5, X_2 = 0.5, X_3 = 1.0)$

$= p(\ X_1 \mid \mu_1, \sigma_1^2\ )p(\ X_2 \mid \mu_2, \sigma_2^2\ )p(\ X_3 \mid X_1, X_2, w_0, w_1, w_2, \sigma_3^2\ )$

$= \mathcal{N}(\ X_1 = 1.5 \mid \mu_1 = 22.63\ , \sigma_1^2 = 0.5831\ ) \times \mathcal{N}(\ X_1 = 0.5 \mid \mu_1 = 1.1045\ , \sigma_1^2 = 0.867\ ) \times$

$\mathcal{N}(\ X_3 = 1.0 \mid -0.0502 + (0.5661)(1.5) + (0.7683)(0.5), 2.5639\ )=$
$\mathcal{N}(\ X_3 = 1.0 \mid 1.1831, \mathbf{2.5639}\ )$

$= \frac{1}{\sqrt{2\pi(0.5831)}} \exp\left\{-0.5\ \frac{(1.5-22.63)^2}{0.5831}\right\} +$
$\frac{1}{\sqrt{2\pi(0.867)}} \exp\left\{-0.5\ \frac{(0.5-1.1045)^2}{0.867}\right\}+\frac{1}{\sqrt{2\pi(\mathbf{2.5639})}} \exp\left\{-0.5\ \frac{(0.5-1.1831)^2}{\mathbf{2.5639}}\right\} = \mathbf{0.0280}$

--End--