Introduction to AI

Assignment 1

February 21, 2023

1. Write True/False for the following conditional independence statements. Justify clearly your answer by showing active/blocked trails as necessary and appropriate rules for them to be active/blocked. [No coding required for this question. Each sub-question has **2 points**]

   (a) $A \perp G \mid \{F\}$
   False. There is a V-structure between A and G.

   $$A \to B \to D \leftarrow G$$

   Given F and thus implying we know D, it couples A and G. Thus A and G are dependent.

   (b) $A \perp G \mid \{E\}$
   False. There is a common cause structure.

   $$A \leftarrow C \to E \to G$$

   A and E are dependent, and thus by cascade structure A and G are also dependent. But given E, it blocks the active path between A and G. Thus A and G are independent given E.

   (c) $A \perp E \mid \{C\}$
   True. A, C and E have a common cause structure.

   $$A \leftarrow C \to E$$

   Given the common cause C, it decouples A and E. There are no other active paths between A and C, thus they are independent.

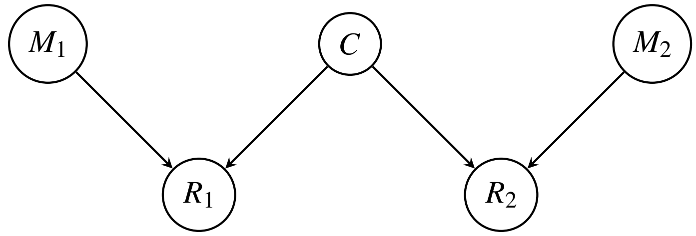
   (d) $A \perp E \mid \{C, D\}$
   False. Similar to part (c), given C it decouples A and E at $A \leftarrow C \to E$. However, there exists a V-structure between A and E.

   $$A \to B \to D \leftarrow E$$

   Similar to part (a), given D it couples A and E. Thus A and E are dependent.

   (e) $A \perp D \mid \{B, E\}$
   True. Cascade from $A \to B \to D$ is blocked given B. Common cause couples A and E. But path from $A \leftarrow C \to E \to D$ is blocked given E. Since all paths from A to D are blocked, A and D are independent.

2. (a) Draw the Bayes net corresponding to this setup.



Variable Name	Domain	Interpretation
C	$\{1, 0\}$	The actual health of a person. Either COVID positive 1 or negative 0.
M_1	$\{a, b, c\}$	The manufacturer of the first test kit. Where the company is a, b or c.
M_2	$\{a, b, c\}$	The manufacturer of the second test kit. Where the company is a, b or c.
R_1	$\{1, 0\}$	The result of the first test kit. Either positive 1 or negative 0.
R_2	$\{1, 0\}$	The result of the second test kit. Either positive 1 or negative 0.

This table can be interpreted as three independent events M_1, M_2 and C. The manufacturer of the two test kits received by a person and their actual health are independent, but they will determine the result of the test kit.

(b) Write conditional probabilities (numerical values) associated with each node of this Bayes net. As there are 5 variables, please specify one conditional probability table (CPT) for each variable.

$P(C = 0)$	$P(C = 1)$
0.7	0.3

$P(M_n = a)$	$P(M_n = b)$	$P(M_n = c)$
0.333	0.333	0.333

M_n	C	$P(R_n = 0 \mid M_n, C)$	$P(R_n = 1 \mid M_n, C)$
a	0	0.99	0.01
b	0	0.95	0.05
c	0	0.91	0.09
a	1	0.3	0.7
b	1	0.2	0.8
c	1	0.1	0.9

For values of $n \in \{1, 2\}$ as each person has two test kits.

(c) Are the results of the two tests dependent or independent given the evidence that the Covid Status is known? Justify your answer.

There is a common cause structure $R_1 \leftarrow C \rightarrow R_2$ which couples R_1 and R_2. But given the COVID status C, it decouples R_1 and R_2. Thus they are independent.

(d) Assume you took both tests at home. After being tested twice in a matter of minutes, the first test was positive and the second negative. What is the probability that you actually have COVID19? Show your analytical computations.

Given that the first test was positive and the second test was negative, then

$$
\begin{aligned}
R_1 &= 1 \\
R_2 &= 0
\end{aligned}
$$

We are trying to solve for $P(C = 1 \mid R_1 = 1, R_2 = 0)$. Since R_1 and R_2 are conditionally independent on C,

$$
\begin{aligned}
P(C = 1 \mid R_1 = 1, R_2 = 0) &= \frac{P(R_1 = 1, R_2 = 0 \mid C = 1)P(C = 1)}{P(R_1 = 1, R_2 = 0)} \\
&= \frac{P(R_1 = 1 \mid C = 1)P(R_2 = 0 \mid C = 1)P(C = 1)}{P(R_1 = 1, R_2 = 0)}
\end{aligned}
$$

To find $P(R_1 = 1 \mid C = 1)$ and $P(R_2 = 0 \mid C = 1)$, marginalise over the manufacturer M.

$$
\begin{aligned}
P(R_1 = 1 \mid C = 1) &= \frac{0.7 + 0.8 + 0.9}{(0.1 + 0.2 + 0.3) + (0.7 + 0.8 + 0.9)} \\
&= 0.8 \\
P(R_2 = 0 \mid C = 1) &= 1 - 0.8 \\
&= 0.2
\end{aligned}
$$

Thus we have solved for numerator,

$$
\begin{aligned}
P(R_1 = 1 \mid C = 1)P(R_2 = 0 \mid C = 1)P(C = 1) &= (0.8)(0.2)(0.3) \\
&= 0.048
\end{aligned}
$$

To solve for the denominator, it is the *hard part*. We cannot use marginalisation as there are too many varibales to sum over. We can use variable elimination instead. The joint distribution is

$$
\begin{aligned}
&P(C = c, M_1 = m_1, M_2 = m_2, R_1 = r_1, R_2 = r_2) \\
=&P(C = c)P(M_1 = m_1)P(M_2 = m_2)P(R_1 = r_1 \mid C, M_1)P(R_2 = r_2 \mid C, M_2)
\end{aligned}
$$

Then,

$$
\begin{aligned}
& P(R_1 = 1, R_2 = 0) \\
&= \sum_{C,M_1,M_2} P(C = c)P(M_1 = m_1)P(M_2 = m_2)P(R_1 = 1 \mid C, M_1)P(R_2 = 0 \mid C, M_2) \\
&= \sum_{C,M_1} P(C = c)P(M_1 = m_1)P(R_1 = 1 \mid C, M_1) \underbrace{\sum_{M_2} P(M_2 = m_2)P(R_2 = 0 \mid C, M_2)}_{\tau_1(R_2=0|C)} \\
&= \sum_{C} P(C = c)\tau_1(R_2 = 0 \mid C) \underbrace{\sum_{M_1} (M_1 = m_1)P(R_1 = 1 \mid C, M_1)}_{\tau_2(R_1=1|C)} \\
&= \sum_{C} P(C = c)\tau_1(R_2 = 0 \mid C)\tau_2(R_1 = 1 \mid C)
\end{aligned}
$$

Calculate the case where $C = 0$,

$$
\begin{aligned}
P(R_1 = 1 \mid C = 0) &= \frac{0.01 + 0.05 + 0.09}{(0.01 + 0.05 + 0.09) + (0.99 + 0.95 + 0.91)} \\
&= 0.05 \\
P(R_1 = 0 \mid C = 0) &= 1 - 0.05 \\
&= 0.95
\end{aligned}
$$

Then using both cases,

$$
\begin{aligned}
P(R_1 = 1, R_2 = 0) &= \sum_{C} P(C = c)\tau_1(R_2 = 0 \mid C)\tau_2(R_1 = 1 \mid C) \\
&= (0.7)(0.95)(0.05) + (0.3)(0.8)(0.2) \\
&= 0.08125
\end{aligned}
$$

Hence,

$$
\begin{aligned}
P(C = 1 \mid R_1 = 1, R_2 = 0) &= \frac{P(R_1 = 1 \mid C = 1)P(R_2 = 0 \mid C = 1)P(C = 1)}{P(R_1 = 1, R_2 = 0)} \\
&= \frac{0.048}{0.08125} \\
&= \frac{192}{325} \\
&\approx 0.59
\end{aligned}
$$