Q3. Bayes Nets, D-Separation and Relevance (worth 28/100 marks)

1. **(worth 5 marks)**. Recall that a Bayesian network is a **directed acyclic graph** that represents a joint distribution. Assume you have a Bayesian network that represents the joint distribution over n variables, $X_1, X_2, ...X_n$. What is the **maximum** number of edges this network can contain? Show how it is possible to construct a Bayesian network with this maximum number of edges, and remember that your network *can't contain any cycles*!

   Solution:
   max edges: $n(n-1)/2$. You can make a BN over X1, X2, . . . Xn by assigning an edge between every pair of variables Xi, Xj such that $j > i$. The total number of edges will then be $n - 1 + (n - 2) + ... + 0 = n(n-1)/2$. If there were a cycle in this graph there would be some path between Xi, Xi+1, Xi+2 .... and back to Xi. But that's impossible because we made the graph by inserting edges between Xi, Xj pairs such that $j > i$, meaning every directed path will see $i < i + 1 < i + 2$ .... and so on. This can't terminate back at $i = j$, because $i < i + 1 < i + 2.... < i$ is impossible.

2. **(worth 5 marks)**. Say we have some Bayesian network with n variables that is able to faithfully represent a distribution $P(X_1, X_2, ...X_n)$. Show that, if X_1 and X_2 are any two variables that are not adjacent to one another, one of two things must be true: either X_1 is conditionally independent of X_2 given the parents of X_1, or X_1 is conditionally independent X_2 given the parents of X_2.

   Solution:
   Consider two cases: (i) X_1 is a descendant of X_2 and (ii) X_1 is not a descendant of X_2. By the definition of d-separation, in the first case X_1 is conditionally independent X_2 given the parents of X_1 and in the second X_1 is conditionally independent X_2 given the parents of X_2

3. **(worth 6 marks; 3 marks each)** For each question below, draw a Bayesian Network on the right hand side that represents the given distribution to the left using **three** arrows to join the variables in the network.

   - $P(A, B, C, D)$ where:

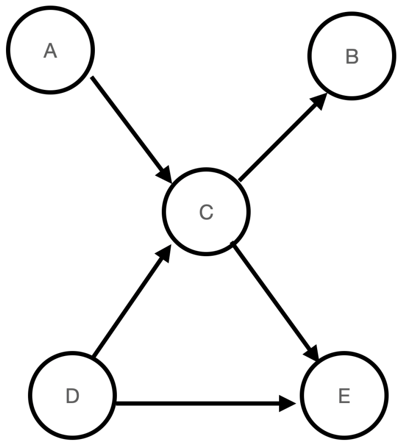
     A **is** independent of B
     A **is** independent of D
     B **is not** independent of D given C
     D **is not** independent of C

     Solutions: $A- > C < -B- > D$ and Solution: $A- > C < -B$ (and also D points to C)

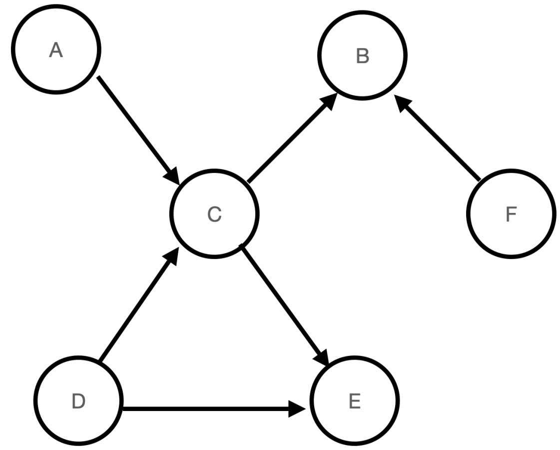
   - $P(A, B, C, D)$ where:

     A **is** independent of C
     A **is not** independent of B
     A **is** independent of B given D
     C **is** independent of B given D

     Solution: $A- > D < -C$ and also D points to B ... maybe there are others?



(a) Version 1



(b) Version 2

Next, consider the Bayesian network structures above. Assume all of the variables in the network are Boolean.

4. **(worth 1 mark)** How many parameters (or table entries in CPTs) are required to fully specify the network in Figure (a) (Version 1)?

   Solution: 12 (or 24)

5. **(worth 1 mark)** List **one set** of variable pairs ($\{X, Y\}$) that are independent in the network described by Figure (a) (Version 1).

   Solution: A and D

6. **(worth 1 mark)** List **two sets** of variable pairs ($\{X, Y\}$) that are independent in the network described by Figure (b) (Version 2).

   Solution: A and D, F and (anything but B)

7. **(worth 1 mark)** List **two sets** of three variables ($\{X, Y, Z\}$) where X is conditionally independent of Z given Y in the network described by Figure (a) (Version 1).

   Solution: A C B and D C B (also E C B)

8. **(worth 3 marks)** Show the steps you would use to calculate $P(D|B = true)$ using variable elimination (VE) using the network described by Figure (a) (Version 1). Use the elimination order A, E, C and avoid unnecessary calculations.

   Solution:
   1. Restrict B (eliminate rows where B=False in $P(B|C)$ to create $F1(C)$)
   2. Eliminate A (multiply$P(A) * P(C|A, D)$ and sum over A to create factor $F2(C, D)$)
   3. Eliminate E (take $P(E|C, D)$ and sum over E to create factor F(C,D)) $<-$ this will be a table of 1s, cuz E is irrelevant, should skip this step!!!!
   4. Eliminate C (sum $F2(C, D) * F1(C)$ over C to create factor F3(D))
   5. Normalize $P(D) * F3(D)$; this is $P(D|B = true)$

9. **(worth 3 marks)** Again, show the steps you would use to calculate $P(D|B = true)$ using variable elimination (VE) in the same network. This time use the elimination order C, E, A.

   Solution:
   1. Restrict B (eliminate rows where B=False in $P(B|C)$ to create $F1(C)$)
   2. Eliminate C (multiply $P(C|D, A) * F1(C)$ and sum over A to create factor F1(D,A))
   3. Eliminate E (take $P(E|C, D)$ and sum over E to create factor F(C,D)) $<-$ this will be a table of 1s, E is irrelevant, should be omitted!!!
   4. Eliminate A (multiply $F1(D, A) * P(A)$ and sum over A to create factor F2(D))
   5. Normalize P(D)*F2(D); this is $P(D|B = true)$

10. **(worth 1 mark)** Which elimination order results in less VE complexity (A, E, C or C, E, A), and **why**?

    Solution: Neither is preferable, both create factors with 2 variables in scope.

11. **(worth 1 mark)** If you had to answer questions (8) and (9) above using using the network described by Figure (b) (Version 2), would your calculations change? In a sentence or less, explain your answer. You don't actually have to perform the calculations, tho!

    Solution: Yes, because F is relevant to the query.

Q4. First-order Logic and Resolution (worth 40/100 marks)

Consider a first-order language L that consists of constant symbols (or 0-ary functions) $o1, o2, o3$, a binary relation symbol $scarier\_than$, and predicate symbol $mammal$. Now say we have a Knowledge Base that consists of the following sentences:

- $mammal(o1)$
- $mammal(o2)$
- $scarier\_than(o2, o1)$
- $scarier\_than(o3, o2)$

Consider a model (or interpretation) $M =< D, \phi, \psi, V >$ such that:

- $D = \{Rabbit, Koala, Alligator\}$
- $\psi(mammal) = \{Rabbit, Koala\}$
- $\psi(scarier\_than) = \{\langle Koala, Rabbit\rangle, \langle Rabbit, Alligator\rangle, \langle Alligator, Koala\rangle\}$
- $\phi(o_1) = Rabbit$
- $\phi(o_2) = Koala$
- $\phi(o_3) = Alligator$

1. **(worth 3 marks)** Does the model M satisfy the Knowledge Base? **Explain** why or why not.
   Solutions: Yes.

   The first and second sentences are satisfied because $\phi(o_1) = Rabbit, \phi(o_2) = Koala$ and $\psi(mammal) = \{Rabbit, Koala\}$.

   The third sentence is satisfied because $\phi(o_1) = Rabbit, \phi(o_2) = Koala$ and $\langle Koala, Rabbit\rangle \in \psi(scarier\_than)$.

   The last sentence is satisfied because $\phi(o_1) = Rabbit, \phi(o_2) = Koala$ and $\langle Alligator, Koala\rangle \in \psi(scarier\_than)$.

2. **(worth 4 marks)** Suppose we add the following sentence to the Knowledge Base:

$$\forall x \forall y \forall z((scarier\_than(x, y) \land scarier\_than(y, z)) \rightarrow scarier\_than(x, z))$$

Does the model M satisfy the Knowledge Base? **Explain** why or why not.

Solutions: No. Consider for example $\mathcal{M}_1 \in \mathfrak{B}$ where
$\langle Koala, Rabbit \rangle \in \psi(scarier\_than)$
$\langle Alligator, Koala \rangle \in \psi(scarier\_than)$
but $\langle Alligator, Rabbit \rangle \notin \psi(scarier\_than)$