**Birla Institute of Technology & Science, Pilani**
**Work Integrated Learning Programmes Division**
**Second Semester 2021-2022**

**Mid-Semester Test**
**(EC-2 Regular)**

| | |
|---|---|
| Course No. | : DSECLZG526 |
| Course Title | : Probabilistic Graphical Models |
| Nature of Exam | : Open Book |
| Weightage | : 30% (As per Course Handout) |
| Duration | : 2 Hours |
| Date of Exam | : 09/07/2022 (AN) |

No. of Pages = 2
No. of Questions = 5

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1. It is known that Asbestos and Smoking are dependent given Cancer. People in Construction have a higher likelihood of being Smokers and being exposed to Asbestos. Being in Construction is independent of having Cancer given exposure to Asbestos and being a Smoker. Draw a Bayesian network that models the relationships between the variables Asbestos, Smoking, Cancer, Construction. Write the joint probability distribution in factored form.

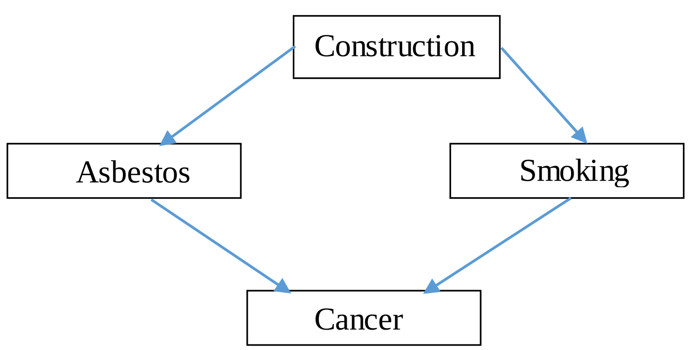
[4 + 2 = 6 Marks]

**Answer**

From the data given in the problem, we see that there must be a converging connection of the form Asbestos → Cancer ← Smoking since Asbestos and Smoking are dependent given Cancer.

We also see that Smoking and Asbestos both depend on Construction, so we have a connection of the form Asbestos ← Construction → Smoking

Taking all of the information given in the problem we have




The joint probability distribution in factored form is

$$P(Construction)\,P(Asbestos/Construction)P(Smoking/Construction)P(Cancer\,/Asbestos,Smoking)$$

Q.2. A Bayesian network on N variables is a complete binary tree, i.e all internal nodes have one edge coming into the node from its parent and exactly two edges leading out into its children. The root node has exactly two edges coming out of it, and each leaf node has exactly one edge leading into it. If each node models a binary random variable, calculate the total number of non-redundant parameters in the network. What is the total number of non-redundant parameters in the unfactored joint probability distribution of all the N variables? Show all steps in your calculation.

[6 Marks]

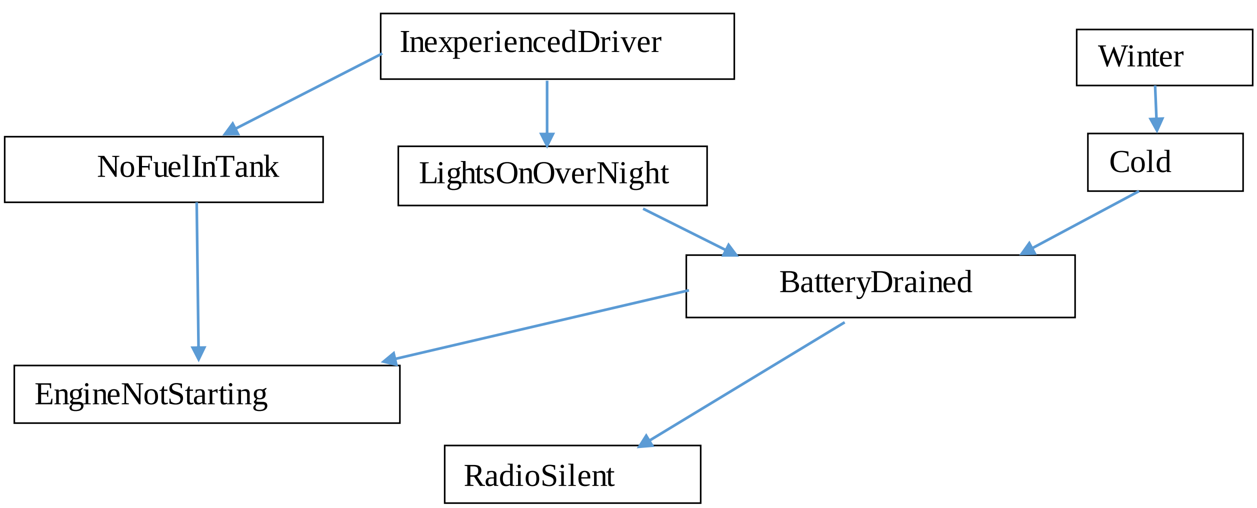
**Answer**

Except for the root node, each node in the network has a single parent and the conditional distribution at each node has the form $P(A/B)$ where $A$ and $B$ are binary random variables. The parameters that need to be preserved at each internal node are $P(A = true/B = true)$ and $P(A = true/B = false)$. The other probabilities at the node are redundant since they can be calculated from the given probabilities. Thus we need 2 non-redundant parameters for each internal node in the network.

For the root node we need only one non-redundant parameter of the form $P(A = true)$. The total number of parameters is therefore $2 * (N - 1) + 1 = 2N - 1$.

The number of non-redundant parameters in the unfactored joint probability distribution is $2^N - 1$.

Q.3. Consider the Bayesian network given below and a few associated CPDs:




| P(LightsOnOverNight/InexperiencedDriver) | InexperiencedDriver = True | InexperiencedDriver=False |
|---|---|---|
| LightsOnOverNight = True | 0.3 | 0.02 |

| P(NoFuelInTank/InexperiencedDriver) | InexperiencedDriver = True | InexperiencedDriver = False |
|---|---|---|
| NoFuelInTank = True | 0.1 | 0.01 |

| P(BatteryDrained = True \| Cold, LightsOnOverNight) | Cold = True | Cold = False |
|---|---|---|
| LightsOnOverNight = True | 0.9 | 0.8 |
| LightsOnOverNight = False | 0.2 | 0.01 |

| P(EngineNotStarting = True \| BatteryDrained, NoFuelInTank) | BatteryDrained= True | BatteryDrained = False |
|---|---|---|
| NoFuelInTank = True | 0.9 | 0.8 |
| NoFuelInTank = False | 0.7 | 0.05 |

(a) Compute P(BatteryDrained=True | InexperiencedDriver=True, Cold=False) with the given data if possible. Otherwise, explain why it is not possible.
(b) Identify the Markov blanket for the node BatteryDrained.

[4 + 2 = 6 Marks]

**Answer**

(a) Let us use the following short-form: B = BatteryDrained, I = InexperiencedDriver and C= Cold and L = LightsOnOverNight. We are asked to compute $P(B/I,\neg C)$. We have

$$P(B,I,\neg C) = P(B,L,I,\neg C) + P(B,\neg L,I,\neg C)$$

$$P(B,I,\neg C) = P(L,I,\neg C) * P(B/L,I,\neg C) + P(\neg L,I,\neg C) * P(B/\neg L,I,\neg C)$$

But $P(B/L,I,\neg C) = P(B/L,\neg C)$ since $B$ is conditionally independent of $I$ given $L$.

Thus $$P(B,I,\neg C) = P(L,I,\neg C) * P(B/L,\neg C) + P(\neg L,I,\neg C) * P(B/\neg L,\neg C)$$

$$P(B,I,\neg C) = P(I,\neg C) * P(L/I,\neg C) * P(B/L,\neg C) + P(I,\neg C) * P(\neg L/I,\neg C) * P(B/\neg L,\neg C)$$

$$\frac{P(B,I,\neg C)}{P(I,\neg C)} = P(B/I,\neg C) = P(L/I,\neg C) * P(B/L,\neg C) + P(\neg L/I,\neg C) * P(B/\neg L,\neg C)$$

Finally we have $P(L/I,\neg C) = P(L/I), P(\neg L/I,\neg C) = P(\neg L/I)$ since $L$ is independent of its non-descendants given its parents. So finally, we have
$P(B/I,\neg C) = P(L/I) * P(B/L,\neg C) + P(\neg L/I) * P(B/\neg L,\neg C)$ which can be computed using the data given above as follows:

$$P(B/I,\neg C) = 0.3 * 0.8 + 0.7 * 0.01 = 0.247$$

(b) The Markov blanket of a node consists of its parents, children, and parents of its children. Therefore the Markov Blanket for $BatteryDrained =$
$\{EngineNotStarting, RadioSilent, LightsOnOverNight, Cold, NoFuelInTank\}$

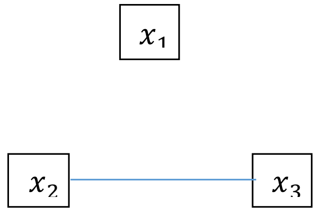
Q.4. Let $P(x_1,x_2,x_3) = f(x_1)g(x_2,x_3)$ be a positive distribution. List all the independencies associated with this distribution, i.e the elements in $I(P)$ with justification. Draw a Markov network to represent the minimal I-map of P.

[4 + 2 = 6 Marks]

**Answer**

From the distribution we note that we note that $x_1$ is independent of $x_2$ and $x_3$.
We have $I(P) = \{x_1 \perp x_2, x_1 \perp x_3, x_1 \perp x_2|x_3, x_1 \perp x_3|x_2\}$ .

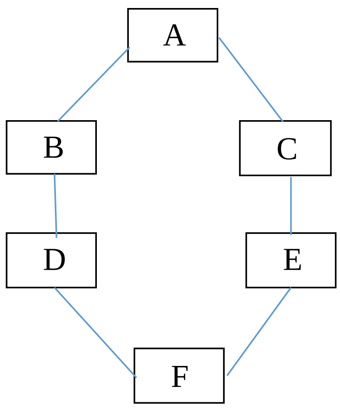
A Markov network to represent the minimal I-map of P is the following:




We construct the above graph using the notion of pairwise independencies – if an edge between $x_i$ and $x_j$ does not exist in the graph it means $x_i$ is independent of $x_j$ given all the other nodes in the graph.

Q.5. Consider the Markov network below, and find a minimal Bayesian network I-map for it using the ordering A,F,B,E,C,D. Provide justification for your answer.

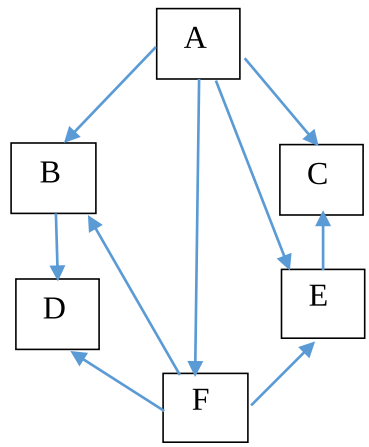
[6 Marks]




**Answer**

The given ordering requires that $A$ be considered as a parent of F, so we draw an edge between $A$ and $F$. The node $B$ must consider $A$ and $F$ as possible parents. $A$ is unavoidable as there is a direct edge to $B$. $F$ must also be considered as a parent of $B$ since $F$ is not independent of $B$ given $A$. For the node $E$, $F$ and $A$ can be considered as parents, and $B$ is independent of $E$ given $F$ and $A$. For node $C$, we consider its parents as $E$ and $A$ since $F$ and $B$ are both independent of $C$ given $E$ and $A$.Similarly for node $D$ we consider $B$ and $F$ as parents and nodes $A, C, E$ are all independent of $D$ given $B$ and $F$.

Thus we add edges $AF, AE, FB$ to the graph and reorient the other edges to obtain the following Bayesian network:


A
B
C
D
E
F