Assignment 2

1 Bayesian networks must be acyclic. Suppose we have a graph $G = (V, E)$ and discrete random variables $X_1 \ldots X_n$ and define

$$f(x_1 \ldots x_n) = \prod_{v \in V} f_n(x_v \mid x_{pa(v)})$$

Where $pa(v)$ refers to the parents of variable X_v in G and $f_v(x_v \mid x_{pa(v)})$ specifies a distribution over X_v for every assignment to X_v's parents i.e. $0 \leq f_v(x_v \mid x_{pa(v)}) \leq 1$ for all $x_v \in Vals(X_v)$ and $\sum_{x_v \in Vals(X_v)} f_v(x_v \mid x_{pa(v)}) = 1$. Recall

that this is precisely the definition of the joint probability distribution associated with the bayesian network G, where the f_v are the conditional probability distributions.

We are to show that is G has a directed ~~graph~~ cycle, f may no longer define a valid probability distribution. Give an example of a cyclic graph G and distributions f_v such that $\sum_{x_1 \ldots x_n} f(x_1 \ldots x_n) \neq 1$. (A valid probability distribution must be non-negative and sum to one) This is why bayesian networks must be defined on acyclic graphs.

→ Solution :-

If there's a cycle in a bayesian network, a variable can be its own ancestor, meaning a node can influence itself through a series of directed edges. This cyclicity can lead to a scenario wher a variable's probability depends on itself, leading to ambiguity.

To illustrate, consider two nodes X_1, X_2 Such that they form a cycle :-

$X_1 \to X_2$
$X_2 \to X_1$

Given the cylce, we have :-

① f_{X_1} - The probability distribution of X_1 (Since X_1 is also a child X_2 due to the cycle).

② $f_{X_2|X_1}$ - The conditional distribution of X_2 given X_1

③ $f_{X_1|X_2}$ - The conditional distribution of X_1 given X_2

Assume,

$f_{x_1} = 0.5$ (probability of x_1 being true)

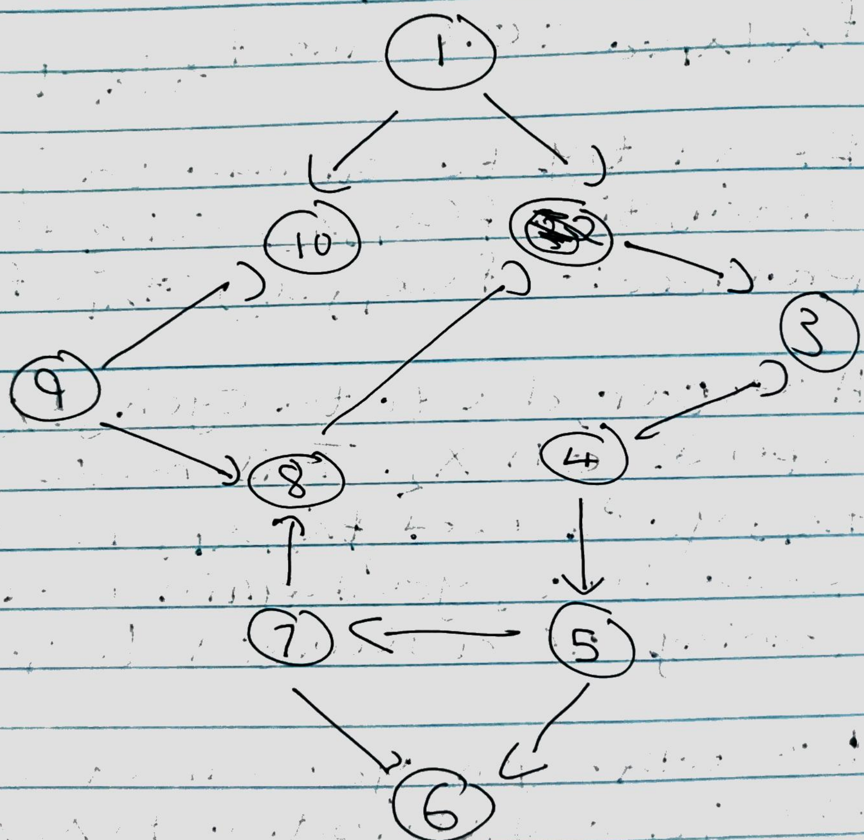
$f_{x_2|x_1=0} = 0.6$ and $f_{x_2|x_1=1} = 0.4$

If we try to calculate the joint distribution $f(x_1, x_2)$ using the formula provided:- $f(x_1, x_2) = f_{x_1} \times f_{x_2|x_1}$

However, due to the cycle, f_{x_1} also depends on x_2. This means to know f_{x_1}, you'd need $f_{x_2|x_1}$, and also know that, you'd again need $f_{x_1|x_2}$ leading to a recursive loop.

Due to the cycle you cannot define the joint distribution $f(x_1, x_2)$ in a way that all possible combinations of x_1 and x_2 sum to 1. This inconsistency arises from the cyclicity in the bayesian network.

Q(2) D-Seperation. Consider the Bayesian networks shown in the below figure



(a) For what pair (i,j) does the statement $X_i \perp X_j$ hold? List all such pairs. And Explain all pairs for $X_i = 4$ or $X_j = 4$.

For d-seperation, we have three rules to determine whether a path between two nodes is blocked

① If we have a chain structure, $A - B - C$ then B blocks the path unless B is observed (conditioned upon).

② For a fork structure $A \leftarrow B \rightarrow C$, B blocks the path unless B or one of its descendants is observed.

Based on graph,

→ 1 and 3 are connected by a chain through 2

→ 1 and 2 are connected by a direct link

→ 2 and 3 are connected by a direct link

→ 3 and 4 are connected by a chain

→ 4 and 5 are connected by a chain through 7

→ 5 and 6 are connected by a direct link

→ 5 and 7 are connected by a direct link

→ 4 and 8 are connected by a direct link

→ 2 and 9 are connected by a direct link

→ 1 and 10 are connected by a direct link.

Thus, the pairs that are not independent are (1, 2), (1, 3), (2, 3), (3, 4), (4, 5), (5, 6), (5, 7)

(4,8), (2,9) and (1,10).

(b) Suppose that we condition on $\{X_2, X_7\}$, shown shaded in the graph. What is the largest set A for which the statement $X_1 \perp X_A \mid \{X_2, X_7\}$ holds? Use the d-seperation algorithm.

→ As we know,

① Conditioning on a node in a chain or a fork blocks information flow.

② Conditioning on a node in a collider unblocks information flow, unless we also condition on a descendant of the collider.

Given conditioning on X_2 and X_7.

X_1 and X_2 are the independent since X_2 is conditioned upon.

X_1 and X_3 Any path from X_1 ~~and~~ to X_3 goes through X_2, which is conditioned. Thus X_1 and X_3 are d-seperated and independent given X_2

X_1 and X_4 :- the path from X_1 to X_4 goes through collider X_3. Normally this would block the phat but since we are conditioned on a child of X_3 (ie X_2) the path is active and X_1 and X_4 are dependent.

X_1 and X_5 :- The path goes through X_4 and X_7. Since X_7 is conditioned upon, this path is blocked, making X_1 and X_5 independent given the conditions.

X_1 and X_6 :- The path goes through X_5 and then X_7, which is conditioned upon. Thus, X_1 and X_6 are d-seperated and independent given X_7.

X_1 and X_8 :- Any path from X_1 to X_8 goes through X_4. Since X_4 and X_1 are dependent (as established earlier), X_1 and X_8 are also dependent

X_1 and X_9 :- The direct path goes through X_2 which is conditioned upon. Thus

X_1 and X_4 are d-seperated and independent given X_2

X_1 and X_{10} these nodes have a direct connection, making them dependent.

$\Rightarrow$ given the conditioning on X_2 and X_7 is the largest set A for which $X_1 \perp A \mid X_2, X_7$ nodes is $A = X_3, X_5, X_6, X_9$

Q③ Consider the following distribution over 3 binary variables X, Y, Z:

$$Pr(x,y,z) = \begin{cases} 1/12 & x \oplus y \oplus z = 0 \\ 1/6 & x \oplus y \oplus z = 1 \end{cases}$$

Where $\oplus$ denotes the XOR function. (make sure your calculation of XOR function is correct by verifying $\sum_{x,y,z} Pr(x,y,z) = 1$.) Show that there is no non-empty directed acyclic graph G such that $I_{d-sep}(G) = I(Pr)$. ($I_{d\text{-}sep}(G)$ refers to the ~~set~~ set of conditional independences induced by the directed graphical structure, as obtained by the d-seperation method).

$\longrightarrow$

(1) Verifying the XOR

$X=0, Y=0, Z=0$ $\quad 0 \oplus 0 \oplus 0 = 0$ $\cancel{P=1/12}$, $P=1/12$

$X=0, y=0, z=1$ $\quad 0 \oplus 0 \oplus 1 = 1$, $P=1/6$

$x=0, y=1, z=0$ $\quad 0 \oplus 1 \oplus 0 = 1$, $P=1/6$

$x=0, y=1, z=1$ $\quad 0 \oplus 1 \oplus 1 = 0$; $P=1/12$

$x=1, y=0, z=0$ $\quad 1 \oplus 0 \oplus 0 = 1$, $P=1/6$

$x=1, y=0, z=1$ $\quad 1 \oplus 0 \oplus 1 = 0$, $P=1/6$

$x=1, y=1, z=0$ $\quad 1 \oplus 1 \oplus 0 = 0$; $P=1/12$

$x=1, y=1, z=1$, $\quad 1 \oplus 1 \oplus 1 = 1$, $P=1/6$

$\Sigma = 1$, so the XOR is correct.

(2) Show there's no DAG G such that $I_{d-sep}(G) = I(P_r)$

We have to show theres no DAG. The XOR function is symmetric; the influence of X, Y, Z on the result is equal. Hence no variable can be shown to be conditionally independent of another given the third.

For any structure of the graph:

→ If its $X \to Z \leftarrow Y$, this suggests Z is influenced by both X and Y. But this doesnt capture the XOR relationship property.

→ For any other possible arrangement, we'll have one of the variables as either a parent or a child, implying conditional independence in some scenario. XOR doesnt allow for such independence.

Conclusion, No graph G can capture the conditional independencies presented by XOR function.

Q4 Markov Property: Let $G = (V, E)$ be an undirected graph and let $X = \{X_1, \ldots X_n\}$ be a collection of random variables defined on its nodes. Recall that a probability distribution over X is globally markov with respect to G if for any disjoint subsets A, B, C such that B seperates A from C in the graph, then

the statement $X_A \perp X_C \mid X_B$ is satisfied.
Now consider two other notions of markovianity. A distribution is pairwise markov wrt to G if for any pair of nodes α, β not directly linked by an edge $e \in E$, the corresponding variables X_α and X_β are conditionally independent given all ~~the~~ remaining variables,

$$\forall_{\alpha,\beta} \in V ; (\alpha,\beta) \notin E, X_\alpha \perp X_\beta \mid X_{V \setminus \{\alpha,\beta\}}$$

A distribution is locally markov wrt G if for any α, a variable X_α when conditioned on all of its neighbors, is independent of the remaining variable.

$$\forall_\alpha , X_\alpha \perp X_{V \setminus (N(\alpha) \cup \{\alpha\})} \mid X_{N(\alpha)}$$

(a) If a distribution is global Markov, then it is locally markov :- True
→ By definition, when a distribution is globally markov, any node X_α is independent of all nodes outside its

its neighborhood.

(b) If a distribution is locally markov, then it is globally markov :- False
→ Being locally markov only ensures indepence when considering immediate neighbors. It does not guarentee the more general conditional independences required by the global markov definition

(c) If a distribution is locally markov then it is pairwise markov :- True
→ When a distribution is locally Markov each variable is conditionally independent of all other variables given its neighbors. This implies that any pair of non-adjacent nodes are conditionally independent given all other nodes, which matches the pairwise markov.

(d) If a distribution is pairwise Markov wrt G, then is locally Markov :- true.
→ By the definition of pairwise Markov, any two non-adjacent variable are conditionally independent given all other

variables, this naturally includes the neighbors of each variables, so the distribution is also locally markov.

(e) If a distribution is positive, then all three markov properties are equivalent: True

→ For positive distribution (where no conditional probabilities are zero), the three definitions are equivalent. This is a well-known property in the realm of graphical models. The non-zero condition allows the three properties to be interchangeable.

Q8 Markov Blankets. Let (G, P) be a bayesian networks. That is $G = (V, E)$ is a DAG and P a distribution that factorizes over G. For a node X, defines its Markov Blanket $MB_G(X)$ as the set of X's parents, its children and the parents of its children. (i.e $MB_G(x) = P_a(X) \cup Ch(X) \cup (P_a(Ch(X)) \setminus \{X\})$.

(a) Show that for any variable X, if we define $W = V \setminus \{\{X\} \cup MB_G(X)\}$, then X and W are d-seperated given $MB_G(X)$.

→ By definition, the markov Blanket, $MB_G(x)$, of a node X includes :-

1. The parents of X.
2. The children of X.
3. The parents of the children of X (other than X itself).

given $MB_G(x)$:

1. X is independent of its parents because it is conditioned on them.
2. X is independent of the parents of its children, since it is seperated from them by its children (on which we condition).
3. X is independent of all nodes not in its markov blanket since none of these nodes have direct or indirect paths to X

Hence, X and W are d-seperated given $MB_G(X)$.

(b) Show that removing any node from the Markov blanket $MB_G(X)$ violates the property you proved in the previous item. That is, for $U \in MB_G(X)$ show that X and $W \cup \{U\}$ are not d-seperated given $MB_G(X) \setminus \{U\}$.

→ (1) ~~If~~ If U is a parent of X :- By removing U, we open up a backdoor path from X to U since they are directly connected

(2) If U is a child of X :- Removing U opens up a path from X to the parents of U (other than X itself) through the child U.

(3) If U is a parent of a child Y of X :- without conditioning on U, there's an open path from X to U through the child Y

In all these cases, X and $W \cup \{U\}$ are not d-seperated given $MB_G(X) \setminus \{U\}$.

(c) Describe a construction that given G, outputs an undirected graph $\hat{G} = (V, \hat{E})$ such that $I(\hat{G}) \subseteq I(G)$ and removing any edge from $\hat{G}$ will violate this i.e. for any $e \in \hat{E}$ define $\hat{G} \setminus \{e\} = (V, \hat{E} \setminus \{e\})$, it holds that $I(\hat{G} \setminus \{e\}) \not\subseteq I(G)$).

→ Construction:

1. Start with G' being a copy of G without the direction on the edges
2. For every v-structure in G (i.e. a structure where two nodes A and B both have directed edges into node C but there's no direct edge between A and B), add an undirected edge between A and B in G'.