Inference and Representation, Fall 2023

Problem Set 2: Directed and Undirected Graphical Models

Due: Monday, Oct 30th, 2023

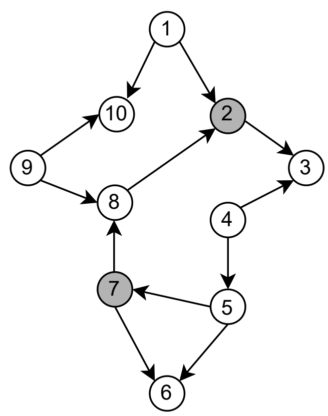
1. **Bayesian networks must be acyclic.** Suppose we have a graph $\mathcal{G} = (V, E)$ and discrete random variables $X_1, \ldots, X_n$, and define

$$f(x_1, \ldots, x_n) = \prod_{v \in V} f_v(x_v | x_{pa(v)}),$$

where $pa(v)$ refers to the parents of variable X_v in $\mathcal{G}$ 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. $0 \le f_v(x_v \mid x_{pa(v)}) \le 1$ for all $x_v \in \mathrm{Vals}(X_v)$ and $\sum_{x_v \in \mathrm{Vals}(X_v)} f_v(x_v \mid x_{pa(v)}) = 1$. Recall that this is precisely the definition of the joint probability distribution associated with the Bayesian network $\mathcal{G}$, where the f_v are the conditional probability distributions.

We are to show that if $\mathcal{G}$ has a directed cycle, f may no longer define a valid probability distribution. Give an example of a cyclic graph $\mathcal{G}$ 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.

2. **D-separation.** Consider the Bayesian network shown in the below figure:



   (a) For what pairs (i, j) does the statement $X_i \perp X_j$ hold? (Do not assume any conditioning in this part, ignore the shading on nodes $2, 7$.) List all such pairs. And explain all pairs for $X_i = 4$ or $X_j = 4$. (To explain one pair $X_i \perp X_j$, list one blocked path between them and tell which local rules are used.)

   (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-separation algorithm.

3. 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 $I_{d-sep}(G) = I(\Pr)$. ($I_{d-sep}(G)$ refers to the set of conditional independences induced by the directed graphical structure, as obtained by the d-separation method).

4. **Markov Property** Let $G = (V, E)$ be an undirected graph and let $\mathcal{X} = \{X_1, \ldots, X_n\}$ be a collection of random variables defined on its nodes. Recall that a probability distribution over $\mathcal{X}$ is *globally Markov* with respect to G if for any disjoint subsets A, B, C such that B separates 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* with respect 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 of 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* with respect to G if for any α, a variable X_α, when conditioned on all of its neighbors, is independent of the remaining variables:
   $$\forall\ \alpha\ ,\ X_\alpha \perp X_{V \setminus (N(\alpha) \cup \{\alpha\})} \mid X_{N(\alpha)}\ .$$
   For each of the following questions, answer *true* or *false*, justifying your answer:

   (a) If a distribution is globally Markov, then it is locally Markov.

   (b) If a distribution is locally Markov, then it is globally Markov.

   (c) If a distribution is locally Markov, then it is pairwise Markov.

   (d) ***Extra credit (5 pts)** If a distribution is pairwise Markov wrt G, then it is locally Markov.

   (e) ***Extra credit (5 pts)** If a distribution is positive, then all three Markov properties are equivalent.

5. **Markov Blankets** Let (G, P) be a Bayesian network. That is, $G = (V, E)$ is a DAG and P a distribution that factorizes over G. For a node X, define its Markov blanket $MB_G(X)$ as the set of X's parents, its children and the parents of its children (i.e. $MB_G(X) = Pa(X) \cup Ch(X) \cup (Pa(Ch(X)) \setminus \{X\})$).

   (a) Show that for any variable X, if we define $\mathbf{W} = V \setminus \{\{X\} \cup MB_G(X)\}$, then X and $\mathbf{W}$ are d-separated 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 $\mathbf{W} \cup \{U\}$ are **not** d-separated given $MB_G(X) \setminus \{U\}$.

   (c) ***Extra credit (5pt)** Describe a construction that given G, outputs an *undirected* graph $\tilde{G} = (V, \tilde{E})$ such that $\mathcal{I}(\tilde{G}) \subseteq \mathcal{I}(G)$ and removing any edge from $\tilde{G}$ will violate this (i.e. for any $e \in \tilde{E}$ define $\tilde{G}_{\setminus\{e\}} = (V, \tilde{E} \setminus \{e\})$, it holds that $\mathcal{I}(\tilde{G}_{\setminus\{e\}}) \not\subseteq \mathcal{I}(G)$).