G-Markov Distributions, D-Separation Theorem

September 9, 2024 *Notes by Jacob Ornelas*

1 Directed Acyclic Graph (DAG)

In order to represent causal relationships and conditional independence, we will use the "Directed Acyclic Graph" (DAG). The Directed Acyclic Graph is made of Vertices: $V = \{1, \ldots, n\}$ (sometimes $\{X_1, \ldots, X_n\}$) and Edges: $E \subseteq V \times V$, where Vertices are connected by Edges, with the condition that there are no self-loops: $(V, V) \notin E, \forall v \in V$. Edges/Paths can either be directed or undirected.

Definition 1 *A* **directed path** $v \to w$ *is a sequence* $v = v_0, v_1, \ldots, v_k = w$ *such that* $(v_i, v_{i+1}) \in E$ *for all* i*.*

Definition 2 *A* **path** v*-*w *is a sequence such that* $\forall i$*, either* $(v_i, v_{i+1}) \in E$ *or* $(v_{i+1}, v_i) \in E$*.*

Example 3 *In this example there are 2 directed paths from a to e: 1) a,b,d,e and 2)a,c,d,e. Note: There are 2 paths from b to c: 1) b,a,c and 2) b,d,c but neither is directed.*

The next few definitions provide some terminology for whether nodes are up-stream or down-stream from one another.

Definition 4 **Parents***:* $par(v) = \{w \in V | (w, v) \in E\}$*.*

Definition 5 **Children***:* $chi(v) = \{w \in V | (v, w) \in E\}$

Definition 6 **Ancestors***:* $anc(v) = \{w \in V | \exists \text{ directed path } w \to v\}$

Definition 7 **Descendants***:* $des(v) = \{w \in V | \exists \text{ directed path } v \to w\}$

Example 8 *Utilizing the structure from Example 1, we observe the following:*

1. $par(d) = \{b, c\}$

2. $chi(d) = \{e\} = des(d)$
3. $anc(d) = \{a, b, c\}$

Definition 9 *G is* **Acyclic** *("DAG") if* $\forall v \in V \nexists$ *directed path* $v \to v$.

Definition 10 **Skeleton***: skel(G) is the* <u>*undirected graph*</u>*:* $(v, \tilde{E})$.
$\tilde{E} \subseteq \binom{V}{2}$ *and* $\tilde{E} = \{\{v, w\} | (v, w) \subseteq E$ *or* $(w, v) \subseteq E\}$.

From here on, we will fix a DAG, G, on $\{1, \ldots, n\}$ with all edges of the form (i,j) for $i < j$.

Definition 11 *Random variables* $X_1, \ldots, X_n$ *are* **G-Markov** *if their joint distribution satisfies*

$$p(x_1, \ldots, x_n) = \prod_{j=1}^{n} p(x_j | (x_i)_{i \in par(j)}).$$

Here, as usual $p(x_1, \ldots, x_n)$ is shorthand for $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$. Informally, what this definition says is that the joint distribution of G-Markov random variables can be

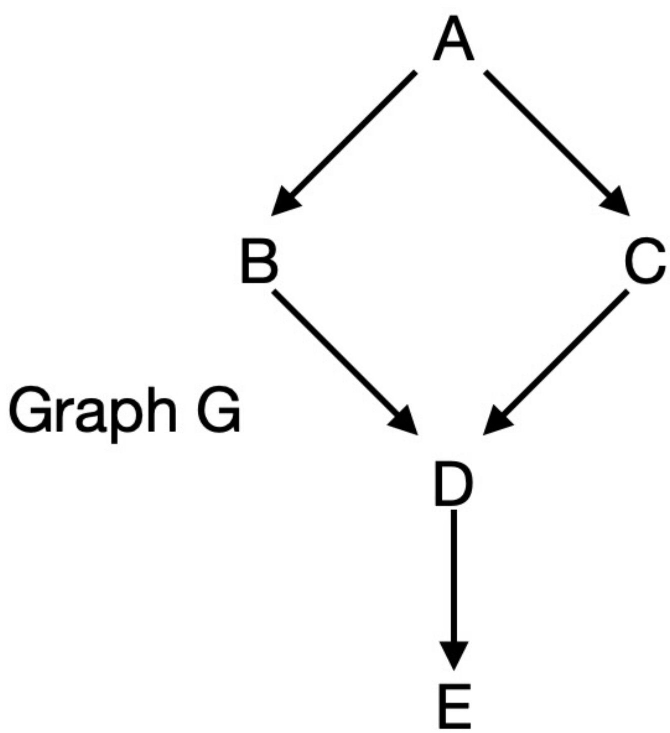

Figure 1: Directed Acyclic Graph: G

decomposed into a product of conditional probabilities where each node is conditioned only on its parents.

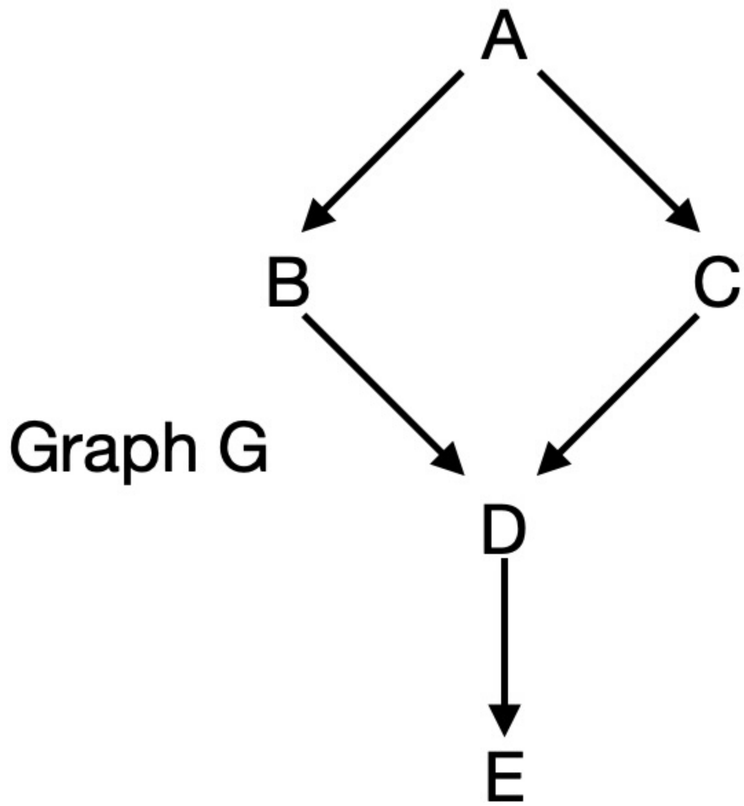

Figure 2: Directed Acyclic Graph: G

Example 12 *Utilizing the structure from Example 1, the joint distribution of G can be written as:* $p(a, b, c, d, e) = p(a)p(b|a)p(c|a)p(d|b, c)p(e|d)$.

Note: Any $X_1, \ldots, X_n$ will be $K_n - Markov$ where $K_n = (\{1, \ldots, n\}, \{(i, j)|i < j\})$ is the complete directed graph on n vertices. This just says that $p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \ldots p(x_n|x_1, \ldots, x_{n-1})$
Goal: Reduce the number of variables we're conditioning on.
Aside: Other Direction Given $X = (X_1, \ldots, X_n)$, find a (small) DAG G such that X is G-Markov. This is called *causal discovery*.
Greedy causal discovery: For fixed ordering of $X_1, \ldots, X_n$, "Markov Parents" of X_j are a minimal subset, $S \subseteq \{1, \ldots, j-1\}$, such that $p(x_j|x_1, \ldots, x_{j-1}) = p(x_j|(x_i)_{i \in S})$. In general these will depend on the ordering.

2 Conditioning can either create or destroy dependence

Given a G-Markov $X_1, \ldots, X_n$ which conditional independence statements $X_i \perp\!\!\!\perp X_j | X_k$ hold?

The next four examples will motivate Pearl's d-Separation Theorem by showing how conditioning can in some cases create dependence and in other cases destroy it!

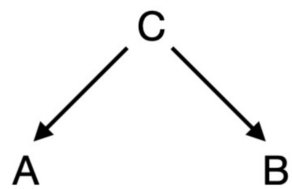

Figure 3: Conditioning on C destroys the dependence between A and B.

Example 13 $p(a, b, c) = p(c)p(a|c)p(b|c)$
$A \not\perp\!\!\!\perp B|\emptyset$: A and B are unconditionally dependent
$A \perp\!\!\!\perp B|C$: A and B are are conditionally independent given C

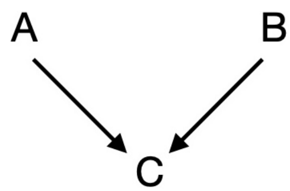

Figure 4: Conditioning on C creates dependence between A and B.

Example 14 $p(a, b, c) = p(a)p(b)p(c|a, b)$
$A \perp\!\!\!\perp B|\emptyset$: A and B are unconditionally independent
$A \not\perp\!\!\!\perp B|C$: A and B are conditionally dependent given C

Example 15

$$A \to C \to B$$

$p(a, b, c) = p(a)p(c|a)p(b|c)$
Is $A \perp\!\!\!\perp B|\emptyset$? No, for example C=A, B=C
Is $A \perp\!\!\!\perp B|C$? Yes
Proof: $p(a, b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)}{p(c)} * p(b|c) = p(a|c)p(b|c)$*, which results from Bayes Rule.*

Example 16 *Is $A \perp\!\!\!\perp B|D$? No, for example D=C*

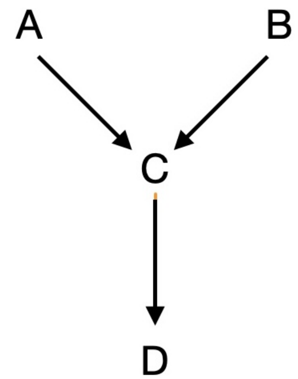

Figure 5: Conditioning on D creates dependence between A and B.

3 The d-Separation Theorem

Let $X_1, \ldots, X_n$ be G-Markov and let A, B, C be disjoint subsets of $\{X_1, \ldots, X_n\}$. The d-separation theorem gives a combinatorial condition that's sufficient for conditional independence $A \perp\!\!\!\perp B \,|\, C$. To state it we'll need a definition.

Definition 17 *Given a path $\gamma = (v_0, \ldots, v_k)$ in G with $v_0 \in A$ and $v_k \in B$, and $0 < j < k$, we say that γ is* **blocked** *if there exists $0 < j < k$ such that*

1. *$v_j \in C$ and $v_{j-1} \leftarrow v_j \rightarrow v_{j+1}$; or*
2. *$v_j \in C$ and $v_{j-1} \rightarrow v_j \rightarrow v_{j+1}$; or*
3. *$v_j \in C$ and $v_{j-1} \leftarrow v_j \leftarrow v_{j+1}$; or*
4. *$v_j \notin C$ and $des(v_j) \cap C = \emptyset$ and $v_{j-1} \rightarrow v_j \leftarrow v_{j+1}$.*

Definition 18 *A and B are* **d-separated** *by C in G, if all paths A-B in G are blocked.*

Now that we have an understanding of what it means to block a path and the meaning of d-separation, we can state the d-separation theorem.

Theorem 19 (d-Separation (Verma & Pearl, 1988)) *Let $(X_1, \ldots, X_n)$ be G-Markov, and let A, B, C be disjoint subsets of $\{X_1, \ldots, X_n\}$. If A and B are d-separated by C in G, then $A \perp\!\!\!\perp B \,|\, C$.*

Conversely, If A and B are not d-separated by C in G, then there exists a G-Markov distribution such that $A \not\!\perp\!\!\!\perp B \,|\, C$.

Notation: We denote that A and B are d-separated by C with $(A \perp\!\!\!\perp B|C)_G$

Exercise 20 (Lauritzen, 1990) *Given subsets of vertices A, B, C in a DAG G, form an undirected graph $L(A \cup B \cup C)$ as follows.*

1. *Delete all vertices not in $anc(A \cup B \cup C) \cup A \cup B \cup C$.*
2. *$\forall v, w \in V$ such that $chi(v) \cap chi(w) \neq \emptyset$, add and edge $\{v, w\} \in E(L)$.*
3. *Remove arrows: $\forall (v, w) \in E$ add an edge $\{v, w\} \in E(L)$.*

Claim: *$(A \perp\!\!\!\perp B|C)_G$ if and only if every path A-B in $L(A \cup B \cup C)$ intersects C.*

Unconditional Case: When is $X_i \perp\!\!\!\perp X_j|\emptyset$ in G-Markov $(X_1, \ldots, X_n)$?
Answer: When all paths $X_i - X_j$ are blocked, i.e. every path $X_i - X_j$ has a collider: $v_{j-1} \to v_j \leftarrow v_{j+1}$.

Note that a path with no collider must have the form

$$X_i \leftarrow \cdots \leftarrow Z \to \cdots \to X_j.$$

The node Z is then a common ancestor of X_i and X_j (or $Z = X_i$ or $Z = X_j$).

Write $\overline{anc}(X) = anc(X) \cup \{X\}$ for the set of ancestors of X including itself.

Corollary 21 (No common ancestor implies unconditional independence) *If $\overline{anc}(X_i) \cap \overline{anc}(X_j) = \emptyset$, then $X_i \perp\!\!\!\perp X_j|\emptyset$.*