DGMs: Markov Properties
10708, Fall 2020
Pradeep Ravikumar

1 Global Markov Properties

Consider a DAG $G = (V, E)$. We can — just by inspecting the graph — specify the following set of conditional independence "Markov" properties. In the sequel, we will use nodes $s \in V$ and RVs X_s interachangeably. In the undirected graph case, there was a very natural notion of graph separation that specified the global Markov properties. In the DAG case however, there is a more subtle notion, called d-separation that specified corresponding global Markov properties.

Given such a notion, we can then define the set of global Markov properties as:

Definition 1 *Given a DAG G, the set of global Markov properties is the set of conditional independencies:*

$$\mathbb{I}(G) = \{\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z} \,:\, \textsc{dsep}_G(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})\}.$$

So let us see how to specify whether $\textsc{dsep}_G(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})$ holds for any tuple of disjoint node subsets $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$. To first set up some notation, we say there is a trail $X - X_{v_1} - X_{v_2} - \ldots - X_{v_k} - Y$ in the DAG G between two nodes X and Y iff $(X_i, X_{i+1}) \in E$ or $(X_{i+1}, X_i) \in E$ for $i \in [k-1]$. Our goal is to specify **blocked trails** between any two nodes X and Y given some set of "observed" nodes $\mathbf{Z}$.

Given such a notion, we have the following:

Definition 2 *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be three sets of nodes in a DAG G. We say that $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$, also denoted as $\textsc{dsep}_G(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})$ if all trails are blocked between any node $X \in \mathbf{X}$ and any node $Y \in \mathbf{Y}$ given $\mathbf{Z}$.*

This mirrors the definition of graph separation in undirected graphs, but the notion of a blocked trail in DAGs is a bit more subtle. So it remains to specify what we mean by blocked trails in DAGs. Let us first consider some base cases.

- **Causal Trail:** $X \to Z \to Y$ is blocked iff Z is observed.
  If this is the entire graph, this then corresponds to the d-separation $\textsc{dsep}(X, Y \,|\, Z)$, and consequently to the conditional independence:

  $$X \perp\!\!\!\perp Y \,|\, Z.$$

  Example: The future is independent of the past given the present.

- **Evidential Trail:** $X \leftarrow Z \leftarrow Y$ is blocked iff Z is observed. If this is the entire graph, this then corresponds to the d-separation $\textsc{dsep}(X, Y \mid Z)$, and consequently to the conditional independence:

$$X \perp\!\!\!\perp Y \mid Z.$$

  Example: The future is independent of the past given the present.

- **Common Cause:** $X \leftarrow Z \rightarrow Y$ is blocked iff Z is observed. If this is the entire graph, this then corresponds to the d-separation $\textsc{dsep}(X, Y \mid Z)$, and consequently to the conditional independence:

$$X \perp\!\!\!\perp Y \mid Z.$$

  Example: Let X denote shoe-size, and Y denote gray hair. The two might be marginally dependent, but conditioned on age Z, they do become conditionally independent.

- **Common Effect:** $X \rightarrow Z \leftarrow Y$ is blocked iff neither Z nor any of its descendants $\textsc{desc}_Z$ is observed. If this is the entire graph, this then corresponds to the d-separation $\textsc{dsep}(X, Y \mid Z)$, and consequently to the conditional independence: $X \perp\!\!\!\perp Y$, but not necessarily that $X \not\perp\!\!\!\perp Y \mid Z$. **Example:** Suppose I am to meet a friend at a cafe, and the friend is late. But my friend is never late, so I know there can only possibly be two scenarios. Either aliens have abducted my friend, or perhaps more alarmingly, my watch is not working. Let us denote the binary possibility of friend being late or not by by the RV Z, that of aliens abducting my friend or not by X, and my watch working or not by Y. Now, we can see that X is independent of Y (or else my friends, knowing how well I take care of my watch, might prefer not to be friends for the imminent danger of alien abduction). But clearly $X = 1$ (alien abduction) given $Y = 1$ (watch not working) and $Z = 1$ (friend late) is much lower than the probability of $X = 1$ given $Z = 1$. So X is not independent of Y given Z.

Let us next consider the case of a general trail $X_1 - \ldots X_n$. Such a trail is blocked given some set of nodes $\mathbf{Z}$ in G iff each trail snippet of consecutive three nodes is blocked given $\mathbf{Z}$. Formally, we have:

Definition 3 *We say that a trail $X_1 - \ldots X_n$ is blocked in a DAG G given some set of nodes $\mathbf{Z}$ in G iff*

- ***v-structure:*** *If there is a v-structure via consecutive nodes $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, then neither X_i nor some node in $\textsc{desc}_{X_i}$ are in $\mathbf{Z}$*

- *some other node that is not the common child in a v-structure is in $\mathbf{Z}$*

1.1 Undirected Graph Characterization of d-separation

There is an alternative characterization of d-separation via traditional graph separation on a specially constructed undirected graph. We first recall some graph theoretic notation. Given a DAG $G = (V, E)$, the induced graph over some subset $U \subseteq V$ is the DAG $G[U] = (U, E')$, where $E' = \{(X, Y) \in E(G) \,:\, X, Y \in U\}$. Let $\textsc{ances}_X$ denote the set of ancestors of X, where $Y \in \textsc{ances}_X$ iff if there is a directed path from Y to X.

Definition 4 (Moralization) *We say that an undirected graph $\mathcal{M}[G] = (V, E')$ is a moral graph corresponding to a DAG $G = (V, E)$when $(X, Y) \in E'$ iff*

- *$(X, Y) \in E$ or $(Y, X) \in E$, or*
- *X and Y are parents of the same node in G*

The term "moralization" is due to the ostensible morality of marrying the parents of a node.

Given this setup, we have the following proposition.

Proposition 5 *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be three disjoint subsets of nodes in a DAG G. Let $U = \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$. Let $G[U \cup \textsc{ances}_U]$ be the induced DAG over U and its ancestors, and let $H = \mathcal{M}[G[U \cup \textsc{ances}_U]]$ be its moralized undirected graph. Then,*

$$\textsc{dsep}_G(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z}) \text{ iff } \textsc{sep}_H(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z}).$$

The proposition can be easily seen to hold for the base cases above, with the general case following by induction.

1.2 Soundness

So far, we have seen a characterization of "global Markov properties" associated with a DAG G. But these are just conditional independence statements we "read off" a DAG. The interesting question is whether a DGM distribution P that *factors according to G* satisfies the global Markov properties associated with G.

Proposition 6 *Any distribution P that* factors according to a DAG *G satisfies the global Markov properties associated with G.*

Proof. We will use the moralized ancestral graph characterization of d-separation (global Markov properties) associated with a DAG. As in the earlier proposition, consider three disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and let $U = \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$. Let $G[U \cup \textsc{ances}_U]$ be the induced DAG over U and its ancestors, and let $H = \mathcal{M}[G[U \cup \textsc{ances}_U]]$ be its moralized undirected graph. Let P be any distribution that factors according to the DAG G. Then we define $P_{\mathcal{M}}$ be the distribution specified by a product of factors each of which is simply the conditional distribution — derived from P — of a node in H given its parents. Since H is ancestrally closed (the parents of any node in H also lie in H), all the nodes involved in these conditional distributions lie in H. It can also be seen $P_{\mathcal{M}}$ factors according the UG H since each conditional distribution factor involves a node and its parents, which are fully connected in the moralized graph.

Now consider any global Markov property $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}$ associated with the DAG G. Then, $\textsc{dsep}_G(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})$, and by the earlier proposition, $\textsc{sep}_H(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})$. Since $P_{\mathcal{M}}$ factors according to the UG H, and by the soundness of UG global Markov properties, we have that $P_{\mathcal{M}}$ satisfies the conditional independence that $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}$. Now $P_{\mathcal{M}}[U \cup \textsc{ances}_U] = P[U \cup \textsc{ances}_U]$ since its factors are precisely all the conditional distributions of nodes given their parents. It thus follows that the distribution P satisfies the conditional independence, as desired. □

As with the UG case, it is worth pausing here to reflect on the remarkable result above: any distribution that factors according to the DAG also satisfies the set of all global Markov properties associated with that DAG. We thus obtain a very simple characterization of (at least some of) the *constraints* satisfied by any DGM P associated with a DAG G. The above property is thus also called *soundness* of the global Markov properties associated with a DAG.

The converse holds as well, and unlike the UG case, does so for all distributions (not just positive ones).

Theorem 7 *Any distribution P that satisfies the global Markov properties associated with a DAG G also* factors according to G.

We will be proving this as a by-product of analyzing *local Markov properties* associated with a DAG G.

2 Local Markov Properties

Unlike the UG case, local Markov properties are actually much more naturally specified for DAGs compared to the more subtle characterization of global Markov properties via d-separation.

Let $\textsc{pa}_X$ denote the set of parents of X, and $\textsc{non-desc}_X$ denote the set of variables in $\mathcal{X}$ that are not descendants of X.

Definition 8 *Given a DAG G, we define the set of local independencies as*

$$\mathbb{I}_\ell(G) = \{X_s \perp\!\!\!\perp \textsc{non-desc}_{X_s} \,|\, \textsc{pa}_{X_s} \,:\, \forall X_s \in \mathcal{X}\}.$$

Thus, each node is conditionally independent of its non-descendants given its parents.

It can be easily seen that $I_\ell(G) \subseteq I(G)$, since it holds that $\textsc{dsep}_G(X_s, \textsc{non-desc}_{X_s} \,|\, \textsc{pa}_{X_s})$. But it turns out that unlike the UG case, the set of distributions satisfying these are the same, even in the general case. We have the following proposition.

Proposition 9 *Given a DAG G, and any distribution P, the following statements are equivalent.*

- *(F) P factorizes with respect to DAG G*
- *(GM) P satisfies global Markov cond. independencies in $I(G)$*
- *(LM) P satisfies local Markov cond. independencies in $I_\ell(G)$*

Proof. We have already shown that F implies GM, and that GM implies LM. So it remains to show that LM implies F. We show this by induction. Let X_s be a leaf of the DAG (i.e. a node with no children). Then by LM, we have that:

$$P(X_s \,|\, X_{-s}) = P(X_s \,|\, \textsc{pa}_{X_s}).$$

By induction, $P(X_{-s})$ factorizes with respect to $G[V - \{X_s\}]$. Then, since $P(X) = P(X_s \,|\, X_{-s})\, P(X_{-s})$, it thus follows that P factorizes with respect to the DAG G. $\square$

We can also define the following set of pairwise conditional independencies, which can also be shown to be equivalent characterization of DGMs.

Definition 10 (Pairwise Markov properties) *Given a DAG G, we define the set of pairwise Markov independencies as the set:*

$$\mathbb{I}_p(G) = \{X \perp\!\!\!\perp Y \,|\, \textsc{non-desc}_X - \{Y\} \,:\, (X,Y) \notin E(G),\, Y \in \textsc{non-desc}_X\}.$$

But these are far less commonly studied than the local or global Markov properties in the case of DAGs, in large part because in any case they are equivalent characterizations of DGMs.

2.1 Completeness

Note that the previous "soundness" theorem entailed that any distribution that factors according to a DAG G iff it also satisfies the global Markov properties associated with G. Note that the global Markov properties are a set of conditional independencies. They are also associated with a set of conditional dependencies: any conditional independence assertion that does not lie in the set of global Markov properties associated with G.

One question is whether any distribution that factors according to UG G satisfies not just conditional independencies wrt G but also conditional *dependencies*. Formally, if it does not hold that $\textsc{dsep}_G(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, then whether $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}$ would not be satisfied by any distribution P that factors according to G.

We can also state the above via the notion of *faithful* distributions.

Definition 11 *A distribution P is said to be faithful to a DAG G if whenever P satisfies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \,|\, \mathbf{Z}$ it holds that $\textsc{dsep}(\mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z})$.*

Thus a converse of the soundness proposition might ask that any distribution P that factors according to G is also faithful to G.

But this turns out to be too strong, and does not actually hold. Instead, we have the following weaker version of "completeness":

Theorem 12 *Suppose X and Y are not d-separated given $\mathbf{Z}$ in UG G. Then there exists a distribution P that factors according to G, where X and Y are dependent given $\mathbf{Z}$.*

In the case of discrete DGMs, we can strenghten this somewhat.

Proposition 13 Almost all *distributions P that factorize over a DAG G, i.e. other than on a set of measure zero in the space of conditional probability factorizations, are also faithful to the graph G.*