ECE 175B: Probabilistic Reasoning and Graphical Models
Lecture 8: d-Separation and Conditional Independence

Florian Meyer & Ken Kreutz-Delgado

Electrical and Computer Engineering Department
University of California San Diego

JACOBS SCHOOL OF ENGINEERING

Z-Active Path in a DAG

Let D be a DAG, $D = (\mathcal{V}, \mathcal{E})$. For convenience, make the abuse of notation $D = \mathcal{V} = \mathbb{X}$, where $\mathbb{X}$ denotes a set of random variables which are modeled as all the nodes in D. Let X, Y, Z denote disjoint subsets of $D = \mathbb{X}$. Recall that a path is an *undirected* path (i.e., a path in the underlying undirected graph of D)

Definition 1. Z-active Path

A path from a node in X to a node in Y is active relative to Z (or Z-active, or just active) if both of the conditions above are true:

1. *Every converging node (aka collider node) on the path is in Z, or has a descendent which is in Z*
2. *Every other node (i.e., every non-converging node) on the path is not in Z*

Any path between X and Y can be analyzed to determine if it is Z-active

Note that a collider node in the DAG does not need to be a collider node on the path

Z-Active Path – Continued

- A Z-active path likely gives the random variables associated with the two end-point nodes of the path the ability to provide statistical information about each other when conditioned on the statistical situation encoded in Z, and we say that the two nodes are *d-connected*, or *graphically dependent*, given Z.
- However, we can *not* 100% rule out *the possibility* that the random variables associated with these nodes are (probabilistically) conditionally independent given the statistical situation encoded in Z.
- This is because whether or not the two end-point nodes (viewed as random variables) are actually (probabilistically) dependent given Z (viewed as a collection of random variables) depends on the particular numerical values of the distribution.

Z–Blocked Paths and d–Separation of DAG Subsets

Definition 2. Z**–blocked Path**

A path from a node in X to a node in Y is blocked by Z (or Z-blocked, or just blocked) if it is not Z-active.

Heuristically, when the path is Z-blocked, we imagine that Z "completely blocks" the transmission of information between the two end-point node to the other. In this case the random variables associated with the path end-points are 100% *guaranteed to be conditionally independent* given Z.

Definition 3. **d-Separation of** X **and** Y **by** Z

The subsets X and Y of the DAG D are d-separated by the subset Z, denoted by $\langle X|Z|Y\rangle_d$, if every path from a node in X to a node in Y is Z-blocked.

Note that *every* path from X to Y must be considered, not just the paths which pass through Z.

Definition 4. d-Connection of X and Y by Z

The subsets X and Y of the DAG D are d-connected by the subset Z if X and Y are not d-separated by Z. This is denoted by

$$[X|Z|Y]_d = \neg\langle X|Z|Y\rangle_d$$

An equivalent, often more useful, definition of d-connectedness is,

Definition $4'$. d-Connection of X and Y by Z

The subsets X and Y of the DAG D are d-connected by the subset Z if at least one path between X and Y is Z-active.

Thus to determine if two disjoint subsets are d-connected or separated by a mutually disjoint subset Z, sequentially check each path from X to Y to see if it is Z-active. If *even one path* is found to be Z-active then *stop* the testing procedure and declare X and Y to be d-connected. If no path is found to be Z-active (i.e., if all paths are Z-blocked), then declare X and Y to be d-separated by Z.

Notation for Statistical Dependence

The following three expressions are equivalent notational statements that X is probabilistically *dependent* on Y when conditioned on Z given the specified numerical values for a specific distribution p,

$$X \top\!\!\!\top_p Y \,\big|\, Z \;\;\equiv\;\; X \not\perp\!\!\!\perp_p Y \,\big|\, Z \;\;\equiv\;\; \neg\left(X \perp\!\!\!\perp_p Y \,\big|\, Z\right).$$

All three of these expressions are used in the lecture, homework, homework solutions, and class exams. Often the specification values of the distribution is tacit or irrelevant (i.e., the above statement is true for all possible distributions p), in which case we simply write

$$X \top\!\!\!\top Y \,\big|\, Z \;\;\equiv\;\; X \not\perp\!\!\!\perp Y \,\big|\, Z \;\;\equiv\;\; \neg\left(X \perp\!\!\!\perp Y \,\big|\, Z\right).$$

Relationship between Graphical Separation and Probabilistic Independence

THEOREM

Take the disjoint subsets X, Y, and Z to represent collections of random variables which correspond to the nodes in DAG D. Then,

$$\langle X|Z|Y\rangle_d \implies X \perp\!\!\!\perp Y \mid Z \quad \textit{for all distributions } p \textit{ consistent with } D$$
$$[X|Z|Y]_d \implies X \not\perp\!\!\!\perp Y \mid Z \quad \textit{for almost all distributions } p \textit{ consistent with } D$$

The last statement means "for *almost all*, but *not all*, distributions." This is because if one chooses probability specification values for p uniformly at random in the space of such values we will have dependence almost surely; *however* a human can choose "nongeneric" numerical specification values that result in independence.

Graphical Separation and Probabilistic Independence

From modal logic we borrow and modify the two symbols,

$$\Box \;\equiv\; \text{“it is guaranteed that”} \quad \text{and} \quad \Diamond \;\equiv\; \text{“it is almost always true that”}$$

and use them as follows,

$$\Box\left(X \perp\!\!\!\perp Y \,\middle|\, Z\right) \;\equiv\; \text{“}X \perp\!\!\!\perp Y \,\big|\, Z \text{ for } \textit{all} \text{ possible spec. values of } p\text{”}$$
$$\Diamond\left(X \top\!\!\!\top Y \,\middle|\, Z\right) \;\equiv\; \text{“}X \top\!\!\!\top Y \,\big|\, Z \text{ for } \textit{almost all} \text{ possible spec. values of } p\text{”}$$

Then we can write the previous theorem as

THEOREM

Take the disjoint subsets X, Y, and Z to represent collections of random variables which correspond to the nodes in DAG D. Then,

$$\langle X|Z|Y\rangle_d \implies \Box\left(X \perp\!\!\!\perp Y \,\middle|\, Z\right)$$
$$[X|Z|Y]_d \implies \Diamond\left(X \top\!\!\!\top Y \,\middle|\, Z\right)$$

Example – The Collider Junction

The Collider Junction $a \to c \leftarrow b$ is a DAG, call it $\mathcal{D}_1$, that encodes the factorization,

$$p(a, b, c) = p(c \big| a, b) p(a) p(b).$$

The node c is a collider node on the path between a and b in $\mathcal{D}_1$. Simple marginalization over the collider node c yields

$$p(a, b) = p(a) p(b) \iff a \perp\!\!\!\perp b$$

showing the independence of a and b which, of course, is no surprise since this property is manifestly encoded in $\mathcal{D}_1$.

Let $Z = \emptyset$. From the definition of d-separation we have,

$$\begin{aligned} \text{collider node } c \notin Z \implies \langle a \big| Z \big| b \rangle_d &\implies \Box \left(a \perp\!\!\!\perp b \,\big|\, Z \right) \\ &\implies \Box \left(a \perp\!\!\!\perp b \,\big|\, \emptyset \right) \implies \Box \left(a \perp\!\!\!\perp b \right) \end{aligned}$$

as expected.

The Collider Junction – Continued

Now let $Z = \{c\}$. From the definition of d-connection we have,

$$\begin{aligned} \text{collider node } c \in Z &\implies \left[a \middle| Z \middle| b\right]_d \\ &\implies \Diamond\left(a \perp\!\!\!\perp b \,\middle|\, Z\right) = \Diamond\left(a \perp\!\!\!\perp b \,\middle|\, c\right). \end{aligned}$$

Thus we suspect that a and b are *likely* dependent conditioned on c. Note that

$$p(a, b \big| c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a \big| b, c) p(b \big| c) p(c)}{p(c)} = p(a \big| b, c) p(b \big| c),$$

which is generally not equal to $p(a \big| c) p(b \big| c)$ (the condition for conditional independence) *except* for the very special case when the specification values of p yield numerical equality $p(a \big| b, c) = p(a \big| c)$ for all possible values of the random variables a, b, and c. Since this special case is a real possibility, we can *not* claim that conditional independence is *guaranteed*.

Example – The Chain Junction

The Chain Junction $a \to c \to b$ is a DAG, call it $\mathcal{D}_2$, encoding the factorization,

$$p(a, b, c) = p(b \big| c) p(c \big| a) p(a).$$

Marginalization over the mediator node c shows that a and b are likely unconditionally dependent (equivalently, likely dependent when conditioned on $Z = \emptyset$) *except* for the case when very special numerical specification values of p exist. Thus at most we can claim that

$$\lozenge \left(a \top\!\!\top b \right) = \lozenge \left(a \top\!\!\top b \,\big|\, \emptyset \right).$$

Note that this is consisted with our rule for handling noncollider nodes on the path from a to b:

$$\begin{aligned} \text{noncollider node } c \notin Z = \emptyset \implies \left[a \middle| Z \middle| b \right]_d &\implies \lozenge \left(a \top\!\!\top b \,\middle|\, Z \right) \\ &= \lozenge \left(a \top\!\!\top b \,\middle|\, \emptyset \right) = \lozenge \left(a \top\!\!\top b \right). \end{aligned}$$

The Chain Junction – Continued

Now let $Z = \{c\}$. Our rule for handling noncollider nodes yields,

noncollider node $c \in Z \implies \langle a \big| Z \big| b \rangle_d \implies \Box \left(a \perp\!\!\!\perp b \,\middle|\, Z \right) \implies \Box \left(a \perp\!\!\!\perp b \,\middle|\, c \right)$

We can show this directly,

$$p(a, b \big| c) = \frac{p(a, b, c)}{p(c)} = \frac{p(b \big| a, c) p(a \big| c) p(c)}{p(c)} = p(b \big| c) p(a \big| c).$$

Note that *every* specification compatible with the graph $\mathcal{D}_2$ *must* satisfy

$$p(a, b \big| c) = p(a \big| c) p(b \big| c).$$

Example – The Fork Junction

Let DAG $\mathcal{D}_3$ be the fork junction $a \leftarrow c \rightarrow b$. It encodes the factorization,

$$p(a, b, c) = p(b \big| c)p(a \big| c)p(c).$$

By merely dividing by $p(c)$ it is immediately evident that a and b are always independent when conditioned on the forking node c. This fact is consistent with our rule for handing noncollider nodes:

$$\begin{aligned} \text{noncollider node } c \in Z = \{c\} &\implies \langle a \big| Z \big| b \rangle_d \\ &\implies \Box \left(a \perp\!\!\!\perp b \,\middle|\, Z \right) \implies \Box \left(a \perp\!\!\!\perp b \,\middle|\, c \right). \end{aligned}$$

The Fork Junction – Continued

On the other hand, if $Z = \emptyset$ our rule gives,

$$\begin{aligned} \text{noncollider node } c \notin Z \implies \left[a \middle| Z \middle| b\right]_d \implies & \Diamond \left(a \top\!\!\top b \,\middle|\, Z\right) \\ = & \Diamond \left(a \top\!\!\top b \,\middle|\, \emptyset\right) = \Diamond \left(a \top\!\!\top b\right). \end{aligned}$$

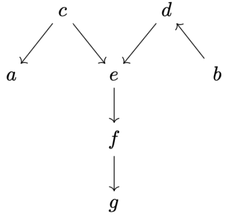
Again we can show this directly by marginalization of c on the factored form for $p(a, b, c)$ shown above, which will generally not lead to the form

$$p(a, b) = \phi_1(a)\phi_2(b)$$

except for very special numerical specification values of $p(a, b, c)$. Because we cannot out of hand preclude such special specification values from occurring, we at best can only make the claim that a and b are *likely* dependent, $\Diamond \left(a \top\!\!\top b\right)$.

Example - Collider and Noncollider Nodes

Consider the path from node a to node b in the following DAG



If the collider node e, or any of its descendants, are in Z, then e is "active" allowing for the potential for "statistical information flow" between nodes a and b. Furthermore, if the noncollider nodes c and d are "left undisturbed", i.e. are not in Z, then they are "active", also allowing for potential "statistical information flow" between nodes a and b. Thus if both of these cases hold it is definitely the case that there likely will be a probabilistic connection (dependency) between the random variables associated with nodes a and b. We say that the path between nodes a and b is active or non-blocked.

Example Continued

$$\begin{tikzcd}
 & c \arrow[dl] \arrow[dr] & & d \arrow[dl] & \\
a & & e \arrow[d] & & b \arrow[ul] \\
 & & f \arrow[d] & & \\
 & & g & &
\end{tikzcd}$$

This is the "Z-active" case and we have,

$$\underbrace{\{e, f, g\} \cap Z \neq \emptyset \text{ and } \{c, d\} \cap Z = \emptyset}_{\text{the path between } a \text{ and } b \text{ is active.}} \stackrel{\text{def}}{\equiv} \underbrace{\left[a \middle| Z \middle| b\right]_d}_{d\text{-connected by } Z} \implies \underbrace{\Diamond \left(a \mathbin{\top\!\!\!\top} b \,\middle|\, Z\right)}_{\text{very likely conditionally dependent}}$$

On the other hand, if the path between nodes a and b is not active (i.e., is blocked), then,

$$\underbrace{\{e, f, g\} \cap Z = \emptyset \text{ or } \{c, d\} \cap Z \neq \emptyset}_{\text{the path between } a \text{ and } b \text{ is blocked.}} \stackrel{\text{def}}{\equiv} \underbrace{\left\langle a \middle| Z \middle| b\right\rangle_d}_{d\text{-separated by } Z} \implies \underbrace{\Box \left(a \perp\!\!\!\perp b \,\middle|\, Z\right)}_{\text{guaranteed conditionally independent}}$$