

3. INDEPENDENCIES

RENATO LUI GEH
NUSP: 8536030

ABSTRACT. This document shows how to represent independencies in bayesian networks. We present ways to find and specify independencies in the network's DAG. We introduce semi-graphoids, graphoids, d-separation and m-separation formally. We then show how to apply d-separation and m-separation on the bayesian network to explore the DAG's variable independency properties. Finally we analyse the BayesBall algorithm proposed by Schachter [Sha98].

1. GRAPHOIDS AND SEMI-GRAPHOIDS

A graphoid follows a set of rules that define variable independence as sets of paths in a graph.

- (1) Symmetry: $X \perp Y|Z \Rightarrow Y \perp X|Z$
- (2) Decomposition: $X \perp Y \cup W|Z \Rightarrow X \perp Y|Z$
- (3) Weak union: $X \perp Y \cup W|Z \Rightarrow X \perp Y|Z \cup W$
- (4) Contraction: $X \perp W|Y \cup Z, X \perp Y|Z \Rightarrow X \perp Y \cup W|Z$
- (5) Intersection: $X \perp Y|Z \cup W, X \perp W|Z \cup Y \Rightarrow X \perp Y \cup W|Z$ if $\mathbb{P} > 0$

A dependency model $I(X, Z, Y) \equiv X \perp Y|Z$ is a graphoid iff it is closed under conditions (i)–(v). A semi-graphoid is closed under (i)–(iv). Rule (v) only holds if the probability function is positive. It was believed that these five rules were axiomatic rules that defined independency accurately [PP85]. However, Studeny later proved that semi-graphoids cannot be used to specify independence exactly [SKV92]. In spite of that, semi-graphoids continue to provide us with a set of useful properties to explore independency in bayesian networks.

2. MARKOV PROPERTIES

In the previous document we had defined Markov Property as:

$$X \perp Nd(X)|Pa(X)$$

This property is also called Local Markov Property (LMP). Another important property that is associated with LMP is the Ordered Markov Property (OMP):

$$X \perp \{Y : Y < X\}|Pa(X), \text{ where } < \text{ is a topological order.}$$

We can use Decomposition, from our previous section, to derive yet another property:

$$X \perp Y | Pa(X) \forall Y \in Nd(X)$$

This property is known as Pairwise Markov Property (PMP). We can derive PMP with Decomposition, yet the other way around (through Composition) does not hold in general unless the distribution is positive, in which case we may apply Intersection and the property holds.

3. TRAILS, PATHS AND CONNECTIVITY OF A GRAPH

Let $G = (V, E)$ be a DAG. A trail is a sequence of nodes X_1, \dots, X_n in V such that (X_i, X_{i+1}) mutual exclusive or (X_{i+1}, X_i) is an arc in E and appears only once. An arc is a directed edge in E that connects two nodes (X, Y) such that $X \rightarrow Y$. The length of a trail is the number of nodes a trail passes on. A directed path is a trail that follows the direction of the arcs. A node X is connected with another node Y if there exists a trail X, \dots, Y . If there is no such trail, then X and Y are separated. A directed cycle is a trail that starts and ends on the same node.

4. D-SEPARATION

A dependency graph is an acyclic and directed graph where its edges are dependencies. Take Figure 1 as example: B depends on A . This means that $A \not\perp B$. A dependency graph must follow the properties defined for graphoids.



FIGURE 1. A dependency graph where nodes X and Y are probabilistically dependent. In this case we have that $X \perp Y$.

D-separation can be summarized into three categories in a dependency graph. The first case happens when we have the following DAG:

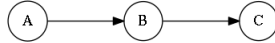


FIGURE 2. In a serial connection we have that $A \perp C | B$ and $A \not\perp C | \emptyset$.

This is called a *serial connection*. Node B blocks nodes A and C , meaning A and C can only be independent of each other (i.e. no dependence edge) if B is removed. Thus, if we can say that $A \perp C | B$ and $A \not\perp C | \emptyset$. That is, if we remove B , we have *unblocked* A and C that were otherwise blocked by B . We say in this situation that A and C are *d-separated* by B .

The second case is known as a *divergent connection*. It happens when a node has two edges that directly connect two distinct other nodes. This can be read as B being the common cause to the effects A and B .

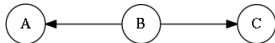


FIGURE 3. In a divergent connection we have that $A \perp C|B$ and $A \not\perp C|\emptyset$.

Note that if we take into account what we have said earlier about serial connectivity, we can apply the same concepts we used and derive the same conclusions. If we remove node B , we have that A and C are now unblocked and are now independent $A \perp C|\emptyset$. Otherwise B blocks A and C and so $A \perp C|B$. Therefore we can say that A and C are d-separated by B .

The third and last case is called the *convergent connection*. In this case we have B as the common effect of A and C (or A and C are the common causes of B). This scenario is known as *explaining away*.

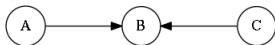


FIGURE 4. In a convergent connection we have that $A \not\perp C|B$ and $A \perp C|\emptyset$.

In this case we have that the trail from A to C is unblocked when B is not removed and blocked otherwise. Node A is independent of C when we have no common effects, and dependent of each other when they have a common effect. We say, in this case, that A and C are d-connected by B . This might seem confusing at first, but take the following situation as example:

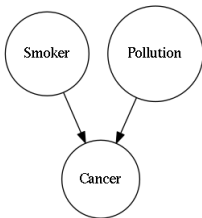


FIGURE 5. *Smoker* and *Pollution* are initially independent of each other. However, if we know our patient has *Cancer*, we might alter the probabilities of our causes accordingly.

Consider the situation where we have a dependency graph that attributes *Smoker* and *Pollution* as common causes to the effect *Cancer*. One could say that smoking and pollution are two independent events (if we assume that smoking does not lead to pollution or that people that smoke are drawn to polluted places), and that smoking and pollution cause cancer (again, we simplify our world to fit our

example). But suppose that our patient has cancer, then we can say that the probability of him smoking or being in constant contact with pollution increases, since they are both possible causes of cancer. If we then discover that our patient also smokes, this might lead to the decrease of our belief that he has been in contact with pollution, since we have found a more probable cause to his cancer. Notice that even though the two causes are independent of each other separately (i.e. if we do not consider cancer), they have an direct influence on each other when one finds that the common effect is known.

Let us now formally define what a blocked and unblocked trail is.

Definition. A trail from X to Y is blocked by a set of nodes \mathcal{B} if:

- (1) X or Y are in \mathcal{B} , or
- (2) There exists a serial or divergent connection X, Y, Z and $Y \in \mathcal{B}$, or
- (3) There exists a convergent connection $X \rightarrow Y \leftarrow Z$ and $(\{Y\} \cup Nd(Y)) \not\subseteq \mathcal{B}$, that is, neither Y nor any of its descendants are in \mathcal{B} .

Otherwise, the trail is unblocked.

Now we define d-separation:

Definition. Sets of nodes \mathbf{X} and \mathbf{Y} are d-separated by a set of nodes \mathbf{Z} if every trail from a variable $X \in \mathbf{X}$ to a variable $Y \in \mathbf{Y}$ is blocked by Z . Otherwise, \mathbf{X} and \mathbf{Y} are d-connected. We will denote d-separation with the symbol \perp_d . Thus, if \mathbf{X} and \mathbf{Y} are d-separated by \mathbf{Z} :

$$\mathbf{X} \perp_d \mathbf{Y} | \mathbf{Z}$$

An alternative way of denoting d-separation is through a function D that takes three arguments $\mathbf{X}, \mathbf{Z}, \mathbf{Y}$. Sets \mathbf{X} and \mathbf{Y} are d-separated from \mathbf{Z} if $D(X, Z, Y)$ is true. D-separation is consistent with LMP because of the following proposition:

Proposition 1. Any node X is d-separated from its nondescendants by its parents.

Our objective is to equate d-separation with independency. For this we must require that d-separation be both sound and complete.

- (1) **Soundness:** $\mathbf{X} \perp_d \mathbf{Y} | \mathbf{X} \Rightarrow \mathbf{X} \perp \mathbf{Y} | \mathbf{Z}$
- (2) **Completeness:** $\mathbf{X} \perp \mathbf{Y} | \mathbf{X} \Rightarrow \mathbf{X} \perp_d \mathbf{Y} | \mathbf{X}$

We know that (1) holds.

Theorem 1. D-separation is sound.

Condition (1) is known as Global Markov Property (GMP). GMP implies LMP and LMP implies GMP. In fact, we can take the following table as a good rule of thumb:

$$\begin{aligned}
GMP &\Rightarrow LMP \Rightarrow PMP \\
PMP &\not\Rightarrow LMP \Rightarrow GMP
\end{aligned}$$

However, d-separation is not complete for all distributions. We will now show a couple of important theorems:

Theorem 2. *For any G graph, there exists a Bayesian network with graph G and variables X and Y such that $X \perp Y$ and $X \not\perp_d Y$.*

Theorem 3. *If G is a DAG and X and Y are d-connected by Z in G , then there is a Bayesian network (G, \mathbb{P}) such that X and Y are dependent given Z .*

5. MARKOV'S BLANKET

A Markov Blanket is defined as:

Definition. *The Markov blanket of a variable X in a graph $G = (V, E)$ is the set \mathcal{B} in which*

$$X \perp V \setminus \mathcal{B} \setminus \{X\} | \mathcal{B}$$

Proposition 2. *Exercise 1*

The set formed by the parents, the children and the parents of the children of a variable X is a Markov blanket of X . If the distribution is positive it is minimal.

Proof. We can summarize all possible outcomes a bayesian network may take with the following network:

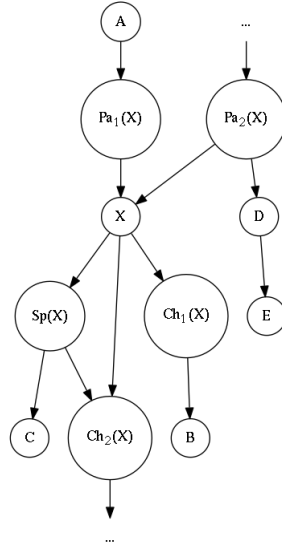


FIGURE 6. All possible connections with X and other nodes outside of the Markov Blanket are either serial or divergent.

$$\begin{aligned}
A &\perp X | Pa(X) \\
B &\perp X | Ch(X) \\
C &\perp X | Sp(X) \\
D &\perp X | Pa(X) \\
E &\perp X | Pa(X)
\end{aligned}$$

The set $Sp(X)$ are the set of spouses of X . That is, $Sp(X) = Pa(Ch(X)) \setminus \{X\}$.

From this bayesian network we can easily see that all possible connections from variable X to or from another node is through a serial or divergent connection. Since a serial or divergent connection implies in $X \perp_d Y \notin \mathcal{B}|\mathcal{B}$, we can extend this, from Theorem 1, that it suffices to show that the markov blanket d-separates X from any other node that is not in \mathcal{B} . \square

D-separation satisfies all properties that define semi-graphoids and also composition and intersection. However, since not all distributions satisfy these two last properties, there are distributions that cannot be fully described by a bayesian network.

6. M-SEPARATION

M-separation verifies independency in an undirected graph.

Definition. A moral graph of a DAG G is the resulting undirected graph of, for every node X , connecting every X to its respective $Sp(X)$ only once.

Definition. Sets of variables \mathbf{X} and \mathbf{Y} are m -separated by \mathbf{Z} if \mathbf{X} and \mathbf{Y} are separated in the moral graph associated with $(An(\mathbf{X}) \cup An(\mathbf{Y}) \cup An(\mathbf{Z})) \setminus \mathbf{Z}$.

Theorem 4. $\mathbf{X} \perp_d \mathbf{Y} | \mathbf{Z}$ iff $\mathbf{X} \perp_m \mathbf{Y} | \mathbf{Z}$, where \perp_m denotes m -separation.

7. BAYESBALL

8. EXERCISES

This section contains the solutions to the proposed exercises from [Dar09].

8.1. Exercise 4.1 (page 72)

a.

$$\begin{array}{ll}
A \perp F, G, H, E | C, D, B & E \perp A, D, C, F | B \\
B \perp C, F, G | D, E, A & F \perp A, B, E | C, D \\
C \perp D, B, E | A & G \perp A, B, C, D, E, H | F \\
D \perp E, C | A, B & F \perp A, B, C, D, G | F, E
\end{array}$$

b.

$$\begin{aligned}
\mathbb{P}(a, b, c, d, e, f, g, h) &= \mathbb{P}(a)\mathbb{P}(b)\mathbb{P}(c|a)\mathbb{P}(d|a, b)\mathbb{P}(e|b)\mathbb{P}(f|c, d)\mathbb{P}(g|f)\mathbb{P}(h|f, e) = \\
&= \theta_a \theta_b \theta_{c|a} \theta_{d|a, b} \theta_{e|b} \theta_{f|c, d} \theta_{g|f} \theta_{h|f, e}
\end{aligned}$$

c.

$$\mathbb{P}(A = 0, B = 0) = \mathbb{P}(\bar{a}, \bar{b})$$

Since we have that A and B are independent of all other nodes, we have that $\mathbb{P}(a, b, c, d, e, f, g, h) = \mathbb{P}(\bar{a}, \bar{b}) = \mathbb{P}(\bar{a})\mathbb{P}(\bar{b})$. Resulting in the following expression:

$$\mathbb{P}(\bar{a}, \bar{b}) = \sum_{c, d, e, f, g, h} \mathbb{P}(\bar{a}, \bar{b}, c, d, e, f, g, h) = \mathbb{P}(\bar{a})\mathbb{P}(\bar{b}) = 0.8 \times 0.3 = 0.24$$

For $\mathbb{P}(e|a)$, since we know that E is independent of A , we can reduce $\mathbb{P}(e|a)$ to the probability $\mathbb{P}(e)$. To find the probability of e , consider the reduced Bayesian network $B \rightarrow E$. $\mathbb{P}(e) = \sum_b \mathbb{P}(e, b) = \sum_b \mathbb{P}(b)\mathbb{P}(e|b)$. Therefore we have that:

$$\mathbb{P}(e|a) = \mathbb{P}(e) = \sum_b \mathbb{P}(b, e) = \sum_b \mathbb{P}(b)\mathbb{P}(e|b) = 1$$

REFERENCES

- [Dar09] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. 1st Edition. Cambridge University Press, 2009.
- [PP85] Judea Pearl and Azaria Paz. “Graphoids: A Graph-Based Logic for Reasoning About Relevance Relations”. In: (1985).
- [Sha98] Ross D. Shachter. “Bayes-Ball: The Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)”. In: *Uncertainty in Artificial Intelligence* (1998).
- [SKV92] M. Studeny, S. Kubik, and J. A. Visek. “Conditional independence relations have no finite complete characterization”. In: (1992).