Chapter 9

Independence in Graphical Models

Graphical models encode independence assumptions. In this chapter, we will study algorithms that enable us to answer questions of the form “Is $S_1 \perp\!\!\!\perp S_2 \mid S_3$?” where S_1, S_2, S_3 are subsets of the nodes in the graph.

Recall that we construct Bayesian network by assuming certain independence assumptions that allow us to remove edges from a complete DAG. The topic of this section is study of all independence properties, which is more general than assumptions used to construct Bayesian networks.

9.1 Independence for sets of random variables

We know that for three random variables x, y, z, x is independent of z given y, denoted $x \perp\!\!\!\perp z \mid y$, if and only if

$$p(x, z \mid y) = p(x|y)p(z|y).$$

This extends to sets of random variables and random vectors. For example, $\{x, y\} \perp\!\!\!\perp \{z, w\} \mid \{t, u\}$, or simply $x, y \perp\!\!\!\perp z, w \mid t, u$, if and only if

$$p(x, y, z, w \mid t, u) = p(x, y \mid t, u)p(z, w \mid t, u)$$

Using this we can show that if $x, y \perp\!\!\!\perp z, w \mid t, u$, then $x \perp\!\!\!\perp z \mid t, u$ and $x, y \perp\!\!\!\perp z \mid t, u$. For example,

$$\begin{aligned} p(x, z \mid t, u) &= \sum_{y', w'} p(x, y', z, w' \mid t, u) \\ &= \sum_{y', w'} p(x, y' \mid t, u)p(z, w' \mid t, u) \\ &= \sum_{y'} p(x, y' \mid t, u) \sum_{w'} p(z, w' \mid t, u) \\ &= p(x \mid t, u)p(z \mid t, u). \end{aligned}$$

Note however that if $x \perp\!\!\!\perp z$ and $y \perp\!\!\!\perp z$ it does not follow that $x, y \perp\!\!\!\perp z$. For a counter-example, set $x \sim \text{Ber}(1/2), y \sim \text{Ber}(1/2)$ and $z = x + y$.

Exercise 9.1. Show that for three disjoint sets of random variables A,B,C, if for some functions f and g,

$$p(A, B \mid C) \propto f(A, C)g(B, C),$$

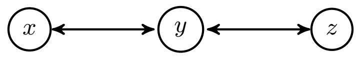
where the constant of proportionality may depend on C, then $A \perp\!\!\!\perp B \mid C$. △

9.2 Independence in Bayesian Networks

In the last chapter we saw that we can obtain a Bayesian Network by starting from a complete graph, representing the chain rule of probability, and then relying on independence assumptions, remove certain edges. Conversely, these independence assumptions are implied by the graphical model. But, in addition, to these, many other independence statements are implied by the network. In this section, we will introduce the concept of *d-separation*, using which we can find all independence statements satisfied by every distribution that factorizes with respect to the Bayesian network. We start by considering several simple networks that will help us describe d-separation.

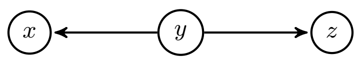
9.2.1 Simple Bayesian networks

Independence analysis in BNs relies on determining when information flows along paths in the graph. As a preliminary step, we study whether information about x affects our belief about z in the graphs of the form given below



with various directions on the edges and with y or one of its descendants being known or unknown.

Example 9.2. Given three random variables x, y, and z with relationships shown below, is $x \perp\!\!\!\perp z$?

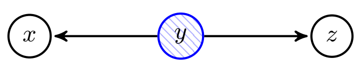


The answer: not in general. The only thing we know from the GM is $p(x, y, z) = p(y)p(x|y)p(z|y)$. We thus have

$$p(x, z) = \sum_y p(x, y, z) = \sum_y p(y)p(x|y)p(z|y)$$

and this is not necessarily equal to $p(x)p(z)$. *Exercise:* Find a counter example, i.e., find p such that it factorizes with respect to the graph but $x \not\perp\!\!\!\perp z$. △

Example 9.3. Is $x \perp\!\!\!\perp z \mid y$ in the graph below?

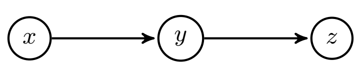


The answer: yes. We need to show $p(x, z \mid y) = p(x|y)p(z|y)$,

$$p(x, z \mid y) = \frac{p(x, y, z)}{p(y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y).$$

△

Example 9.4. Is $x \perp\!\!\!\perp z$ in the graph below?

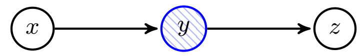


The answer: not in general since

$$p(x, z) = \sum_y p(x)p(y|x)p(z|y) = p(x) \sum_y p(y|x)p(z|y)$$

is not necessarily equal to $p(x)p(z)$. *Exercise:* Provide a counter example for $x \perp\!\!\!\perp z$. △

Example 9.5. Is $x \perp\!\!\!\perp z \mid y$ in the graph below?

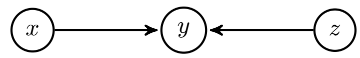


The answer: yes. We have

$$p(x, z \mid y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x)p(y|x)p(z|y)}{p(y)} = \frac{p(y)p(x|y)p(z|y)}{p(y)} = p(x|y)p(z|y).$$

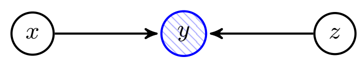
△

Example 9.6. Is $x \perp\!\!\!\perp z$ in the graph below?



Yes: $p(x, z) = \sum_y p(x, y, z) = \sum_y p(x)p(z)p(y \mid x, z) = p(x)p(z) \sum_y p(y \mid x, z) = p(x)p(z)$. △

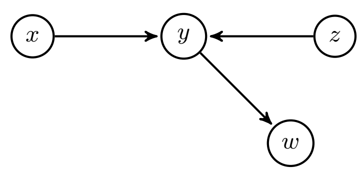
Example 9.7. Is $x \perp\!\!\!\perp z \mid y$ in the graph below?



Not in general. *Exercise:* Verify that for $x \sim \text{Ber}(\frac{1}{2}), z \sim \text{Ber}(\frac{1}{2})$ and $y = x + z$, $p(x, y, z)$ factorizes with respect to the graph above and $x \not\perp\!\!\!\perp z \mid y$. △

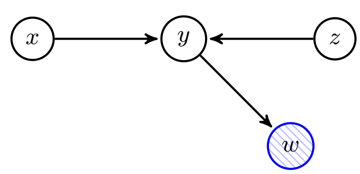
In graphs of Examples 9.6 and 9.7, if y has a descendant, that will also affect the independence relationship between x and z. These cases are considered next.

Example 9.8. Is $x \perp\!\!\!\perp z$ in the graph below?



Yes: $p(x, z) = \sum_{y,w} p(x, y, z, w) = \sum_{y,w} p(x)p(z)p(y \mid x, z)p(w|y) = p(x)p(z) \sum_{y,w} p(y \mid x, z)p(w|y) = p(x)p(z)$. △

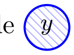
Example 9.9. Is $x \perp\!\!\!\perp z \mid y$ in the graph below?

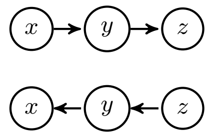
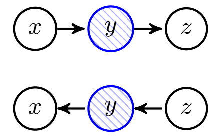
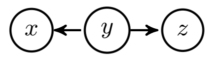
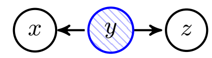
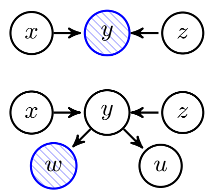
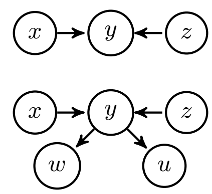


Not in general. Exercise: Verify that for $x \sim \text{Ber}(\frac{1}{2}), z \sim \text{Ber}(\frac{1}{2})$, $y = x+z$, and $w = y$, $p(x, y, z, w)$ factorizes with respect to the graph above and $x \not\perp\!\!\!\perp z \mid y$. △

9.2.2 d-separation

Based on our analysis in the previous section, we can summarize whether information flows from x to z in a graph of the form $x - y - z$ in Table 9.1. The table is organized by the direction of edges at y, with H (Head) representing an incoming edge and T (Tail) representing an outgoing edge. We can see that for the HT, TH, and TT configurations, y blocks the path from x to z if it is known (given) and for HH, it blocks the path if it is not known and neither are any of its descendants.

Table 9.1: Flow of information between x and z. Nodes with style  are knwon.

	Passing through	Blocked
HT/TH		
TT		
HH		

We can generalize this observation to decide, for three disjoint sets A, B, and C, of nodes, whether $A \perp\!\!\!\perp B \mid C$.

Definition 9.10. For a set C of known/observed nodes, a path is said to be blocked if it has a node v such that the nodes incident to v are:

- HT,TH, or TT and $v \in C$;
- HH, and neither v nor its descendants are in C.

Definition 9.11. For disjoint sets A, B, and C, we say that A and B are **d-separated** given C if every path between a node in A and a node in B is blocked if we assume the nodes in C are known.

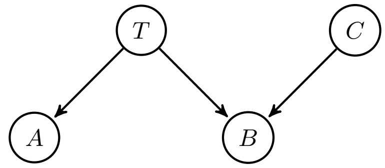
Theorem 9.12. *For three disjoint sets of nodes, A, B, and C, in a graph G, such that A and B are d-separated given C, then $A \perp\!\!\!\perp B \mid C$ according to any probability p that factorize with respect to G.*

Remark ** The converse of the theorem also holds in the sense that any distribution p that satisfies all independencies implied by d-separation factorizes with respect to the graph.

Remark ** Could a distribution p that factorizes with respect to G satisfy independencies that are not implied by d-separation? Indeed, yes. The distribution $p = \prod_{i=1}^{n} p(x_i)$ factorizes with respect to any graph G and for any non-trivial G, p satisfies independencies that are not implied by d-separation in G. However, for any independency $A \perp\!\!\!\perp B \mid C$ not implied by d-separation, there is a probability distribution factorizing with respect to G for which $A \not\perp\!\!\!\perp B \mid C$.

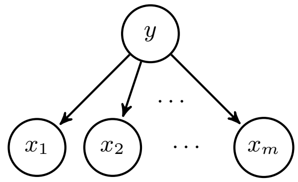
Example 9.13. In the traffic graphic from last chapter, shown below, we want to find all independences of the form $x \perp\!\!\!\perp y$ and $x \perp\!\!\!\perp y \mid z$ for vertices x, y, z. For those that do not follow form d-separation, we write $x \not\perp\!\!\!\perp y$ and $x \not\perp\!\!\!\perp y \mid z$. We have

- No conditioning: $T \perp\!\!\!\perp C$, $T \not\perp\!\!\!\perp A$, $T \not\perp\!\!\!\perp B$, $C \perp\!\!\!\perp A$, $C \not\perp\!\!\!\perp B$, $A \not\perp\!\!\!\perp B$.
- Given T: $A \perp\!\!\!\perp B \mid T$, $A \perp\!\!\!\perp C \mid T$, $B \not\perp\!\!\!\perp C \mid T$.
- Given C: $A \not\perp\!\!\!\perp B \mid C$, $A \not\perp\!\!\!\perp T \mid C$, $B \not\perp\!\!\!\perp T \mid C$.
- Given A: $T \not\perp\!\!\!\perp B \mid A$, $T \perp\!\!\!\perp C \mid A$, $B \not\perp\!\!\!\perp C \mid A$.
- Given B: $T \not\perp\!\!\!\perp A \mid B$, $T \not\perp\!\!\!\perp C \mid B$, $A \not\perp\!\!\!\perp C \mid B$.



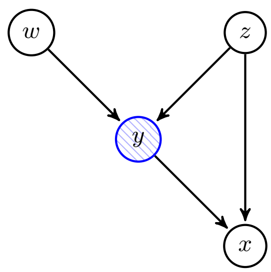
In addition, we have $A \perp\!\!\!\perp \{B, C\} \mid T$ but $\{T, A\} \not\perp\!\!\!\perp C \mid B$. △

Example 9.14 (The Naive Bayes model). The graph for the naive Bayes classification model is



where y denotes the class and $x_1, \ldots, x_m$ denote the dimensions of the input vector. Given y the dimensions are independent, i.e., $x_i \perp\!\!\!\perp x_j \mid y$ for $i \neq j$. But if the class y is not known, generally speaking, $x_i \not\perp\!\!\!\perp x_j$. △

Example 9.15. For four nodes w, x, y, and z, shown below, assume y is given. We can determine that none of the independencies $w \perp\!\!\!\perp z \mid y, x \perp\!\!\!\perp z \mid y, x \perp\!\!\!\perp w \mid y$ follow from d-separation. In fact, we can find a counter example, i.e., a distribution that factorizes with respect to the graph below and does not satisfy these independencies. Specifically, let $w \sim \text{Ber}(1/2), z \sim \text{Ber}(1/2)$, $y = w + z$ and $x = y + z$. Note however that $y \perp\!\!\!\perp n \mid y$ for $n \in \{x, w, z\}$ by the definition of independence.



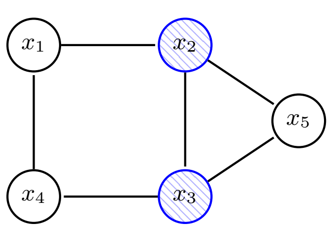
△

9.2.3 Markov Blanket in Bayesian Networks

In a graphical model, the **Markov blanket** of a node y is the set of nodes S such that $y \perp\!\!\!\perp U \mid S$ for any set U. In other words, the set S isolates y from the rest of the graph. In a Bayesian network, the Markov blanket of y consists of its parents, its children, and the immediate parents of its children. The proof of this statement is left as an exercise. An example is shown in Figure 9.1.

9.3 Independence in MRFs

The set of independencies implied by an MRF are more straightforward as separation is the naive graph-theoretic separation. As an example, consider the friendship graph of the previous chapter and assume we know the political affiliation of x_2,x_3.



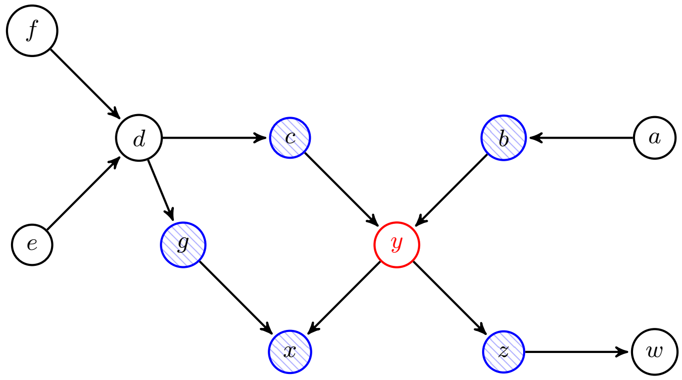


Figure 9.1: The Markov blanket of node y are the set of nodes colored red.

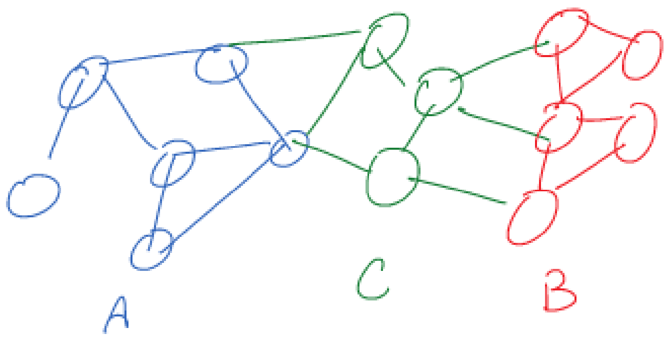


Figure 9.2: MRF Theorem

Then intuitively, we can expect that knowing x_5 does not provide any relevant information about x_1,x_4 and so we must have $x_1, x_4 \perp\!\!\!\perp x_5 \mid x_2, x_3$.

In an MRF G, suppose x_A, x_B, and x_C are disjoint subsets of vertices such that $x_A \cup x_B \cup x_C = G$, as shown in Figure 9.2. *If every path from x_A to x_B travels through x_C, then $x_A \perp\!\!\!\perp x_B \mid x_C$.* To see that this is the case, note that

$$\begin{aligned} p(x_A, x_B \mid x_C) &= P(x_A \mid x_C)P(x_B \mid x_C) \\ &= \frac{p(x_A, x_B, x_C)}{p(x_C)} \propto p(x_A, x_B, x_C) \\ &\propto \prod_{Q \text{ is a clique in } G} \psi_Q(x_Q) = \prod_{Q \in x_A \cup x_C} \psi(x_Q) \prod_{Q \in x_B \cup x_C} \psi(x_Q). \end{aligned}$$

The last equality follows from the fact that there is no clique in G that has a node in both x_A, and x_B since x_C separates x_A and x_B. The result follows from Exercise 9.1.

Examples are given in Figure 9.3.

The **Markov Blanket** of a node in an MRF is the set of neighbors as shown in Figure 9.4.

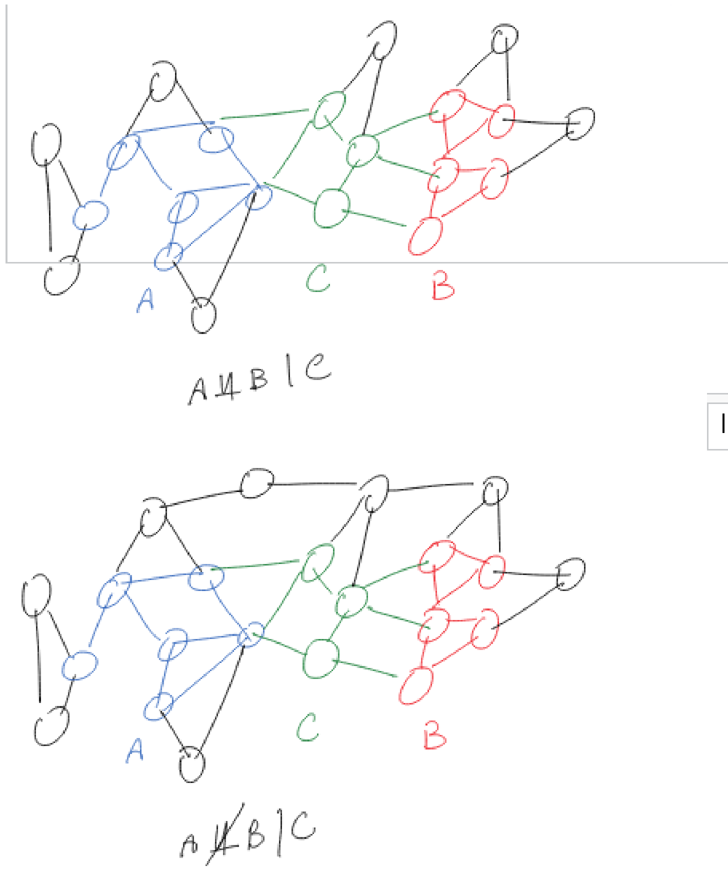


Figure 9.3: Two examples of MRFs

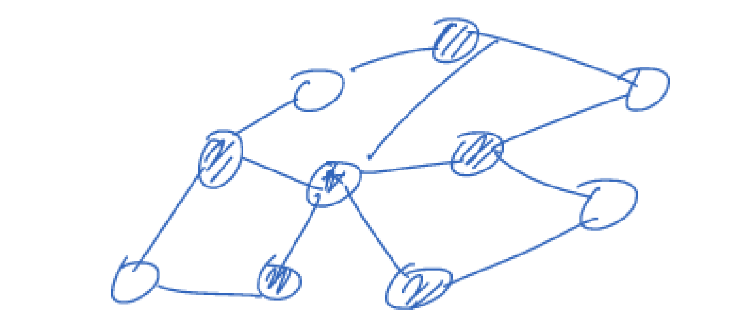


Figure 9.4: As example of a Markov Blanket