Conditional Independence, Dependency-Separation, and Bayesian Networks

Robert Jacobs
Department of Brain & Cognitive Sciences
University of Rochester
Rochester, NY 14627, USA

August 8, 2008

Reference: The material in this note is taken from Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Consider three variables a, b, and c, and suppose that the conditional distribution of a given b and c is such that it does not depend on the value of b [i.e., $p(a|b,c) = p(a|c)$]. In this case, a is conditionally independent of b given c. This can be expressed in a slightly different way by considering the joint distribution of a and b given c, which can be written as follows:

$$p(a,b|c) = p(a|b,c)p(b|c) \quad (1)$$

$$= p(a|c)p(b|c) \quad (2)$$

Note that our definition of conditional independence requires that this equation holds for every possible value of c, and not just for some values.

Consider the Bayesian network in Figure 1. Based on this graph, we can write the joint distribution of a and b given c as follows:

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} \quad (3)$$

$$= p(a|c)p(b|c) \quad (4)$$

meaning that a and b are conditionally independent given c. We can provide a simple graphical interpretation of this result by considering the path from a to b via c. Node c is said to be "tail-to-tail" with respect to this path because the node is connected to the tails of the two arrows. When c is unobserved, the presence of this path causes a and b to be dependent. However, when c is observed and we condition on c, then the conditioned node "blocks" the path from a to b and causes a and b to become conditionally independent.

Now consider the network in Figure 2. Based on this graph, the joint distribution of a and b given c is:

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} \quad (5)$$

$$= p(a)p(c|a)p(b|c) \quad (6)$$

$$= p(a|c)p(b|c) \quad (7)$$

meaning that, again, a and b are conditionally independent given c. The node c is said to be "head-to-tail" with respect to the path from a to b. If c is unobserved, then such a path connects a and b and renders them dependent. But if c is observed, then this observation "blocks" the path, and a and b become conditionally independent.

Lastly, consider the networks in Figure 3. First, we consider the case when node c is unobserved (left graph). In this case a and b are independent. Next, we consider the case when c is observed (right graph). Based on this graph, the joint distribution of a and b given c is:

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} \tag{8}$$

$$= \frac{p(a)p(b)p(c|a,b)}{p(c)} \tag{9}$$

which does not factorize into the product $p(a|c)p(b|c)$ meaning that a and b are not conditionally independent given c. Graphically, we say that node c is "head-to-head" with respect to the path from a to b because it connects to the heads of the two arrows. When c is unobserved, it "blocks" the path, and the variables a and b are independent. However, conditioning on c unblocks the path and renders a and b dependent. There is one more subtlety to this third example. As a matter of terminology, a node y is a descendent of a node x if there is a path from x to y in which each step of the path follows the directions of the arrows. Then it can be shown that a head-to-head path becomes unblocked if either the node or any of its descendants is observed.

In the graphical models literature, people talk about the d-separation property for directed graphs. Consider a general directed graph in which A, B, and C are arbitrary nonintersecting sets of nodes (whose union may be smaller than the complete set of nodes in the graph). We want to know if A and B are conditionally independent given C. To determine this, we consider all possible paths from any node in A to any node in B. Any such path is said to be blocked if it includes a node such that either:

- the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or
- the arrows meet head-to-head at the node, and neither the node nor any of its descendants is in C.

If all paths are blocked, then A is d-separated from B by C.

Consider the graphs in Figure 4. In the left graph, the path from a to b is not blocked by f because it is a tail-to-tail node for this path and it is not observed, nor is it blocked by node e because, although the latter is a head-to-head node, it has a descendant c which is in the conditioning set. Thus, a and b are not conditionally independent given c. In the right graph, the path from a to b is blocked by node f because this is a tail-to-tail node that is observed, meaning that a and b are conditionally independent given f. In addition, this path is also blocked by e because e is a head-to-head node and neither it nor its descendants are in the conditioning set.

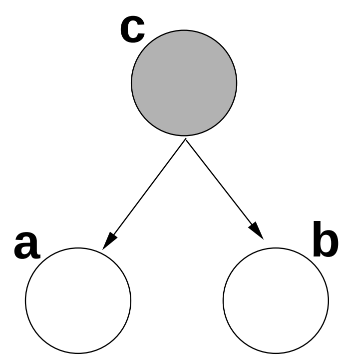

Figure 1: Node c is a parent of a and b. The value of c is observed.

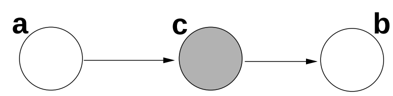

Figure 2: Node a is a parent of c which, in turn, is a parent of b. The value of c is observed.

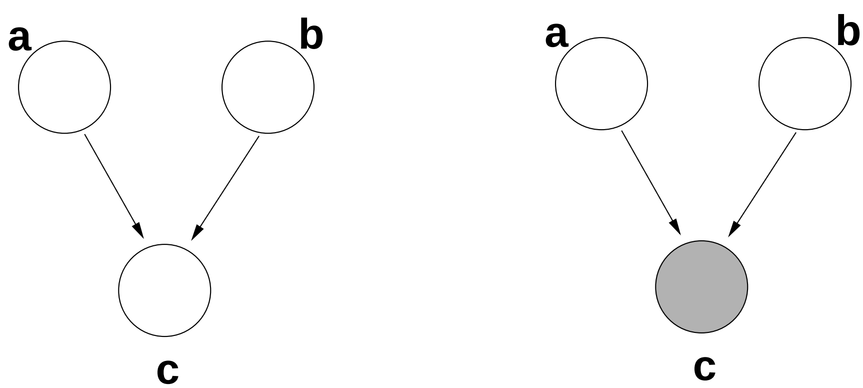

Figure 3: Nodes a and b are parents of c. In the left graph, the value of c is unobserved. In the right graph, it is observed.

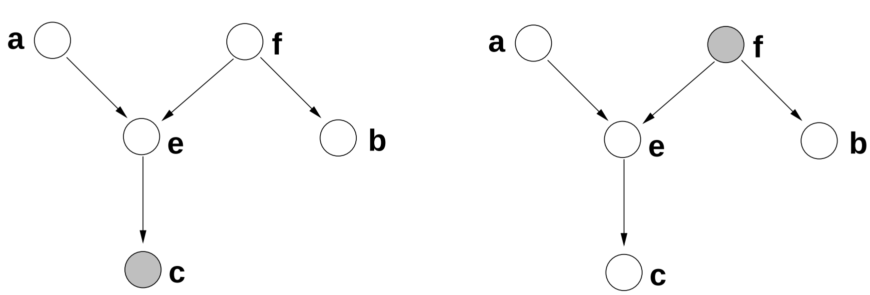

Figure 4: Graphs referred to in text.