2.6 Recovering Latent Structures

When Nature decides to "hide" some variables, the observed distribution $\hat{P}$ need no longer be stable relative to the observable set O. That is, we are no longer guaranteed that, among the minimal latent structures compatible with $\hat{P}$, there exists one that has a DAG structure. Fortunately, rather then having to search through this unbounded space of latent structures, the search can be confined to graphs with finite and well-defined structures. For every latent structure L, there is a dependency-equivalent latent structure (the projection) of L on O in which every unobserved node is a root node with exactly two observed children. We characterize this notion explicitly as follows.

Definition 2.6.1 (Projection)
A latent structure $L_{[O]} = \langle D_{[O]}, O \rangle$ *is a* projection *of another latent structure* L *if and only if:*

1. *every unobservable variable of* $D_{[O]}$ *is a parentless common cause of exactly two non-adjacent observable variables.*

2. *for every stable distribution* P *generated by* L*, there exists a stable distribution* P' *generated by* $L_{[O]}$ *such that* $I(P_{[O]}) = I(P'_{[O]})$.

Theorem 2.6.2 (Verma 1993)
Any latent structure has at least one projection.

It is convenient to represent projections using a bidirectional graph with only the observed variables as vertices (i.e., leaving the hidden variables implicit). Each bidirected link in such a graph represents a common hidden cause of the variables corresponding to the link's endpoints.

Theorem 2.6.2 renders our definition of inferred causation (Definition 2.3.6) operational; it can be shown (Verma 1993) that the existence of a certain link in a distinguished projection of any minimal model of $\hat{P}$ must indicate the existence of a causal path in every minimal model of $\hat{P}$. Thus, our search reduces to finding the distinguished protection of any minimal model of $\hat{P}$ and identifying the appropriate links. Remarkably, these links can be identified by a simple variant of the IC algorithm, here called IC*, that takes a distribution $\hat{P}$ and returns a *marked* pattern, which is a partially directed acyclic graph that contains four types of edges:

1. a marked arrow $a \overset{*}{\rightarrow} b$, signifying a directed path from a to b in the underlying model;

2. an unmarked arrow $a \rightarrow b$, signifying either a directed path from a to b or a latent common cause $a \leftarrow L \rightarrow b$ in the underlying model;

3. a bidirected edge $a \longleftrightarrow b$, signifying a latent common cause $a \leftarrow L \rightarrow b$ in the underlying model; and

4. an undirected edge $a–b$, standing for either $a \leftarrow b$ or $a \rightarrow b$ or $a \leftarrow L \rightarrow b$ in the underlying model.[8]

IC* Algorithm (Inductive Causation with Latent Variables)

Input: $\hat{P}$, a sampled distribution.

Output: $core(\hat{P})$, a marked pattern.

1. For each pair of variables a and b, search for a set S_{ab} such that a and b are independent in $\hat{P}$, conditioned on S_{ab}.
   If there is no such S_{ab}, place an undirected link between the two variables, $a - b$.

2. For each pair of nonadjacent variables a and b with a common neighbor c, check if $c \in S_{ab}$.
   If it is, then continue.
   If it is not, then add arrowheads pointing at c (i.e., $a \rightarrow c \leftarrow b$).

3. In the partially directed graph that results, add (recursively) as many arrowheads as possible, and mark as many edges as possible, according to the following two rules:

[8] Spirtes et al. (1993) used $a \circ\!\!\rightarrow b$ to represent uncertainty about the arrowhead at node a. Several errors in the original proof of IC* were pointed out to us by Peter Spirtes and were corrected in Verma (1993). Alternative proofs of correctness, as well as refinements in the algorithm, are given in Spirtes et al. (1993).

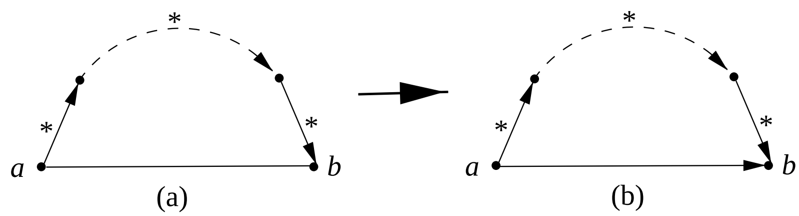

Figure 2.2: Illustration R_2 in step 3 of the IC* Algorithm.

- R_1: For each pair of non-adjacent nodes a and b with a common neighbor c, if the link between a and c has an arrowhead into c and if the link between c and b has no arrowhead into c, then add an arrowhead on the link between c and b pointing at b and mark that link to obtain $c \stackrel{*}{\to} b$.

- R_2: If a and b are adjacent and there is a directed path (composed strictly of marked links) from a to b (as in Figure 2.2), then add an arrowhead pointing toward b on the link between a and b.

Steps 1 and 2 of IC* are identical to those of IC, but the rules in step 3 are different; they do not orient edges but rather add arrowheads to the individual endpoints of the edges, thus accommodating bidirectional edges.

Figure 2.3 illustrates the operation of the IC* Algorithm on the sprinkler example of Figure 1.2. (shown schematically in Figure 2.3(a)).

1. The conditional independencies entailed by this structure can be read off using the d-separation criterion (Definition 1.2.3), and the smallest conditioning sets corresponding to these independencies are given by $S_{ad} = \{b, c\}$, $S_{ae} = \{d\}$, $S_{bc} = \{a\}$, $S_{be} = \{d\}$, and $S_{ce} = \{d\}$. Thus, step 1 of IC* yields the undirected graph of Figure 2.3(b).

2. The triplet (b, d, c) is the only one that satisfies the condition of step 2, since d is not in S_{bc}. Accordingly, we obtain the partially directed graph of Figure 2.3(c).

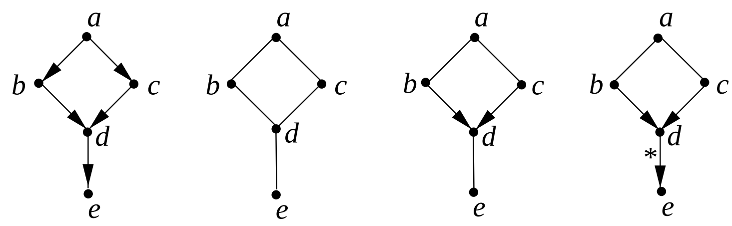

Figure 2.3: Graphs constructed by the IC* Algorithm. (a) Underlying structure. (b) After step 1. (c) After step 2. (d) Output of IC*.

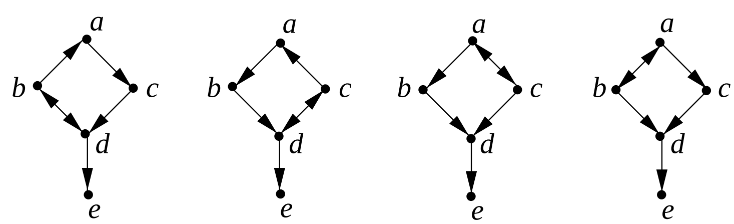

Figure 2.4: Latent structures equivalent to those of Figure 2.3(a).

3. Rule R_1 of step 3 is applicable to the triplet (b, d, e) (and to (c, d, e)), since b and e are nonadjacent and there is an arrowhead at d from b but not from e. We therefore add an arrowhead at e, and mark the link, to obtain Figure 2.3(d). This is also the final output of IC*, because R_1 and R_2 are no longer applicable.

The absence of arrowheads on $a - b$ and $a - c$, and the absence of markings on $b \to d$ and $c \to d$, correctly represent the ambiguities presented by $\hat{P}$. Indeed, each of the latent structures shown in Figure 2.4 is observationally equivalent to that of Figure 2.3(a). Marking the link $d \to e$ in Figure 2.3(d) advertises the existence of a directed link $d \to e$ in each and every latent structure that is independence-equivalent to the one in Figure 2.3(a).