Causal Criterion for Multivariate Correlation under Postselection

Marina Maciel Ansanelli[1,2] **Daniel Centeno**[1,2] **Elie Wolfe**[1] **Matt Jones**[3] **Matthew F. Pusey**[4]

[1]Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5
[2]Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1
[3]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309
[4]Department of Mathematics, University of York, Heslington, York YO10 5DD, United Kingdom

1 ABSTRACT

The field of causal inference examines the relationship between statistical correlations and causal connections among a set of variables. The causal structure is depicted using a directed acyclic graph (DAG), which can include both observed and latent (unobserved) nodes. It is particularly interesting to study if some variables are allowed by the causal structure to show correlation among them, i.e., if one knows the value of one variable, one is able to gain some knowledge about the value of all the others. The interest in this question arises because correlation is useful for different information processing tasks.

In the simplest scenario with two observed variables, it is easy to see that correlation among them is only possible if the corresponding nodes are connected by either a direct causal influence, a latent common cause, or both. These causal structures are represented in Fig. 1. Notice that in both scenarios the two observed nodes share a common ancestor, i.e., there is a node in their common causal past. In the first case, Fig. 1a), the common ancestor is the node A itself (notice that one node can be its own ancestor according to our definition); in the second case, Fig. 1b), it is the latent common cause C and, in the third case, Fig. 1c), both A and C can take that role.

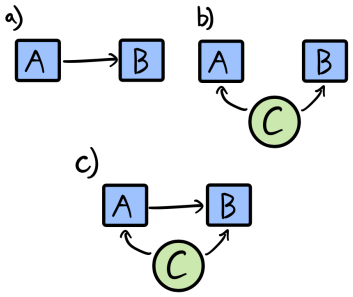

Figure 1: Causal scenarios with two observed nodes with direct influence or a latent common cause.

There is a third way through which correlation between these two variables can be established, which is through postselection (or selection bias). This way requires that the two variables have a common effect, as in Fig. 2. Then, conditioning on the value of the effect, one generates correlation among the initially independent nodes. A simple example to gain intuition about how postselection can help induce correlation is the following. Let A be a variable that indicates if it has rained or not, B a variable associated to the activation of a sprinkler (which does not have any rain detector) and S a variable that shows if the floor is wet or not. Clearly, the sprinkler and the rain are completely independent a priori and both are causes of the floor being wet, so the causal structure is Fig. 2. Then, if one postselects on the value indicating that the floor is wet, one can infer that if it has not rained then the sprinkler was activated, showing that after the postselection some correlations emerge between A and B Pearl and Mackenzie [2018].

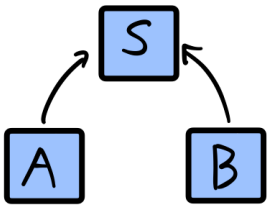

Figure 2: Causal scenario with 3 observable nodes with a common effect.

The most useful graphical tool to address the question of whether some nodes can be correlated at all is the concept of d-separation. This is a ternary relation between three set of nodes, X, Y, Z, that serves as a criterion to answer whether X can be correlated with Y given the value of Z. The concept is to associate stastistical dependence with connection in the causal graph (i.e., the existence of a connecting path) and statistical independence with separation in the causal graph. We can see that in the examples with 2

variables and no postselection given in Fig. 1, A and B are always d-connected given the empty set, i.e., conditioning on nobody.Then, in the example with postselection of Fig. 2, again we check that the 2 variables are d-connected when conditioning on S.

Now, the interesting question is how to know if a set of N arbitrary nodes in a given causal structure can be correlated or not. As a first step, we consider the case of perfect correlation among all the nodes which means that one learns with certainty the value of all the nodes from knowing the value of one of them. The case of not having the possibility of postselecting on any node was studied by Steudel and Ay in Steudel and Ay [2015]. They used the idea of common ancestors and showed that to be able to observe perfect correlation among a set of N variables, there must be a common ancestor shared by all of them.

Then, the remaining question is what is the criterion to state that some set of N nodes is able to be perfectly correlated when one allows for postselection on some variables. In this work, we answer this question.

To understand the answer, first look at the DAG of Fig. 3. There, it is possible to achieve perfect correlation among the nodes A, B, C, D, E, F, G when we condition on S_1 and S_2. This can be done through the following choice of causal parameters:

- The root nodes (i.e., nodes without causal ancestors), A, D, F, G, are completely random bits.
- $B = 0$ if $F = G = 0$; $B = 1$ if $F = G = 1$ and $B = 2$ otherwise
- $S_1 = 0$ if $A = B$ and $S_1 = 1$ otherwise
- $S_2 = 0$ if $B = C$ and $S_2 = 1$ otherwise

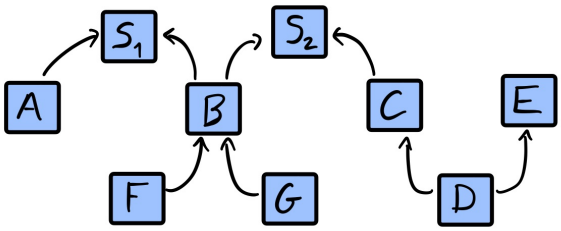

Figure 3: Causal scenario.

This leads us to observe perfect correlation among A, B, C, D, E, F, G when condition on $(S_1, S_2) = (0, 0)$, that is,

$$P(a, b, c, d, e, f, g|s_1 = 0, s_2 = 0) = \frac{1}{2}[0000000]+\frac{1}{2}[1111111]. \quad (1)$$

However, in the DAG of Fig.4, it is impossible to observe perfect correlation among the nodes A, B, C, D, E, F even when we condition on S_1. This is because the node E is d-separated of A, B and C given S_1. Hence, the causal structure imposes that E is statistically independent of A, B, C conditioning on S_1,

$$p(a, b, c, e|s_1) = p(a, b, c|s_1)p(e|s_1). \quad (2)$$

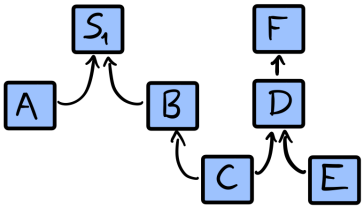

Figure 4: Causal scenario.

With these two examples in mind, we can illustrate the general result that we found in this work: we showed that perfect correlation among the nodes of a set X when conditioning on another set of nodes S is possible if and only if there is a root node R that is d-connected to every node in the set X given the nodes of S. In the example of Fig. 3, all of the root nodes of the DAG satisfy this condition while in the example of Fig. 4, there is no root node that is d-connected to A, B, C, D, E, F given S_1

References

Judea Pearl and Dana Mackenzie. *The Book of Why*. Basic Books, New York, 2018.

Bastian Steudel and Nihat Ay. Information-theoretic inference of common ancestors. *Entropy*, 17(4):2304–2327, 2015.