Everything that can be learned about a causal structure with latent variables by observational and interventional probing schemes

Marina Maciel Ansanelli[1] **Elie Wolfe**[1] **Robert Spekkens**[1]

[1]Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario Canada N2L 2Y5

arXiv link: `https://arxiv.org/abs/2407.01686`

Abstract

When is it impossible to distinguish between two causal structures with latent variables from statistical data obtained by probing each visible variable? If we simply passively observe each visible variable, then it is well-known that many different causal structures can realize the same joint probability distributions. Even for the simplest case of two visible variables, for instance, one cannot distinguish between one variable being a causal parent of the other and the two variables being confounded by a latent common cause. However, it is possible to distinguish between these two causal structures if we have recourse to more powerful probing schemes, such as the possibility of intervening on one of the variables and observing the other. Herein, we address the question of which causal structures remain indistinguishable even given the most informative types of probing schemes on the visible variables. We find that two causal structures remain indistinguishable if and only if they are both associated with the same mDAG structure (as defined in Evans [2015]). We also investigate to what extent one can weaken the probing schemes implemented on the visible variables, such as allowing only for do-interventions that can fix a variable to *one* of its possible values affects, and still have the same discrimination power as a maximally informative probing scheme.

1 SUMMARY OF WORK

The goal of causal discovery is to uncover the causal relations that hold among a set of variables by experimentally probing them. The simplest way through which a variable can be probed is via *passive observation*, that is, by observing and recording the natural value taken by the variable. For instance, we might observe that, in the general population, people that take a specific drug are more prone to recovering from a disease. This does not, however, imply that the drug causes recovery: the observed correlation could be due to a hidden common cause — an unmeasured factor that influences both the likelihood of taking the drug and the likelihood of recovering. For example, an individual's level of health awareness might simultaneously increase the probability of both taking the drug and recovering. Variables such as this one, that causally influence the variables of interest but for whatever reason are not probed, are called *latent variables*. By contrast, the variables that are accessible to be probed are called *visible variables*.

Figure 1 shows the two causal structures that are being adjudicated between in this example: in (a), all of the correlation between drug and recovery is explained by a latent common cause. In (b), such a latent common cause is present, but there is also a direct causal effect.

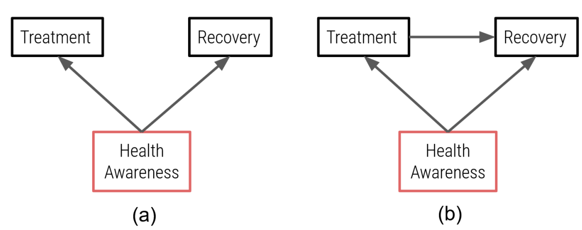

Figure 1: Two causal hypotheses of the example.

To eliminate this ambiguity one could probe the visible variables in a different way, via a *blind drug trial*. In this experiment, each subject is assigned the drug or a placebo at random. This breaks the causal connection between the variable "took the drug or not" to any latent common cause; if a positive correlation persists, it can only be explained by the causal influence of the drug on recovery. In short, by availing oneself of *interventional* probing schemes such as the blind drug trial, one can resolve some ambiguities that appear from the data obtained via passive observations.

In this work, we investigate when such ambiguities remain *regardless* of how the visible variables are probed. The *marginalized DAG* (mDAG) structure, that was introduced in Ref. Evans [2015], emerges as the answer to this question: we showed that two causal structures that are associated with the same mDAG are indistinguishable even when one uses an *informationally complete* probing scheme, that is, one that recovers all the information about the causal structure that is obtainable by interacting with the visible variables [1]. Furthermore, such informationally complete probing schemes can always distinguish causal structures that are associated with different mDAGs. One example of an informationally complete probing scheme is a scheme where, for each visible variable, one observes its natural value and then one implements a do-intervention, which forces the variable to take a fixed desired value. We refer to this as the *Observe&Do (O&D)* probing scheme.[2]

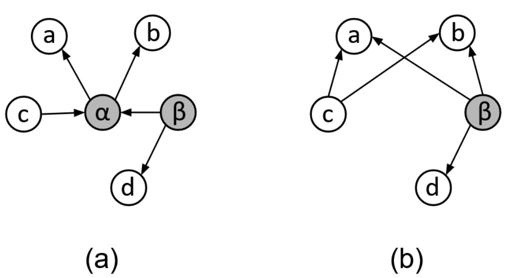

Figure 2: Two causal structures that are associated with the same mDAG. Here, visible nodes are represented in white and latent nodes are represented in gray.

In other words, an informationally complete probing scheme determines the causal structure up to its mDAG equivalence class: two causal structures that are associated with the same mDAG are indistinguishable and two causal structures that are associated with different mDAGs can always be distinguished. Our results therefore establish that the mDAG structure is a fundamental structure for causal analysis. Figure 2 shows an example of two causal structures that are associated with the same mDAG.

In this work, we also investigate to what extent one can weaken the probing scheme and still maintain the discriminatory power of an informationally complete probing scheme. Imagine the case where one is experimentally restricted to, for a given variable, perform *either* a passive observation *or* a do-intervention, without previously observing the natural value of the variable. Further assume that one can do this for all of the visible variables, and that the ensemble of samples is partitioned such that in each partition a different set of visible nodes is intervened upon. The set of visible nodes that is intervened upon in a certain partition of the ensemble (with the complementary set being passively observed) is called a *do-pattern*. If one collects data from do-interventions performed on all possible do-patterns of visible variables, they are implementing what we call the *all-patterns Observe-or-Do (O-or-D) probing scheme*.

The all-patterns O-or-D probing scheme represents a restricted experimental power when compared to the O&D probing scheme, and it is *not* informationally complete. However, as it turns out, two causal structures that correspond to different mDAGs can *also* be distinguished by the all-patterns O-or-D probing scheme.

It turns out that we can restrict our experimental power even more and still be able to distinguish two causal structures that correspond to different mDAGs: it is not necessary to perform do-interventions that set the variables to more than one value. When we are restricted to do-interventions that set the variables to only one value (which we might conventionally call 0) and we partition the ensemble in the same way as described above, we are performing the *all-patterns Observe-or-1Do (O-or-1D) probing scheme*. To see why such a restriction can be of interest, imagine an experiment where we are allowed to force the subjects to quit smoking, but we cannot ethically force them to start smoking. In this case, we can intervene on the experiment to force the variable "smoker" to take the value 0, but not to take the value 1. The all-patterns O-or-1D probing scheme *also* determines the causal structure up to its mDAG equivalence class.

In short, in this work we show that two causal structures that are associated with the same mDAG are indistinguishable even when there is access to an informationally complete probing scheme (such as O&D interventions on all visible variables), which corresponds to a very strong experimental power. Furthermore, we also show that two causal structures that are associated with different mDAGs can always be distinguished by data obtained from the all-patterns O-or-1D probing scheme, which correspond to a much weaker experimental power.

Apart from solving this indistinguishability problem, in this work we also fully characterize when one causal structure *dominates* another relative to the three different types of probing schemes described above, that is, when the first causal structure can realize all the sets of data that are realizable by the second. The characterization of these dominance relations, like the equivalence relations, is consistent across the three probing schemes and is determined by the structure of the corresponding mDAGs.

[1] Edge interventions [Shpitser and Tchetgen, 2014] are not considered among the probing schemes investigated in this work. Nonetheless, we believe that our conclusions can be easily extended to the case where one has access to edge interventions.

[2] Informationally complete probing schemes provide us with the Effect of the Treatment on the Treated (ETT) Shpitser and Pearl [2009].

References

Robin J. Evans. Graphs for margins of bayesian networks. *Scandinavian Journal of Statistics*, 43(3):625–648, November 2015. ISSN 1467-9469. doi: 10.1111/sjos.12194. URL `http://dx.doi.org/10.1111/sjos.12194`.

Ilya Shpitser and Judea Pearl. Effects of treatment on the treated: Identification and generalization. In Jeff A. Bilmes and Andrew Y. Ng, editors, *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, pages 514–521. AUAI Press, 2009. URL `https://www.auai.org/uai2009/papers/UAI2009_0073_63adfb8ce831dbcafdce2ecabcc391fc.pdf`.

Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions, 2014.