Constraint-based Causal Structure Learning when Faithfulness Fails

Jan Lemeire
Vrije Universiteit Brussel
Brussels, Belgium
jan.lemeire@vub.ac.be

Kris Steenhaut
Vrije Universiteit Brussel
Brussels, Belgium
kris.steenhaut@vub.ac.be

Abstract

Constraint-based causal structure learning algorithms rely on the faithfulness property. For faithfulness, all conditional independencies should come from the system's causal structure. The problem is that even for linear Gaussian models the property is not tenable. In this paper, we identify 4 non-causal properties that generate conditional independencies and investigate whether they can be recognized by patterns in the conditional independencies.

1 Introduction

Constraint-based algorithms for learning the causal structure from data rely on quite heavy assumptions, such as faithfulness and the correctness of independence tests. In contrast with the more robust scoring-based algorithms which search for the minimal model among all Bayesian networks, the constraint-based algorithms rely on the conditional independencies that follow from the system's causal structure, which is described by a graph. They can be retrieved from it with the d-separation criterion. If all the conditional independencies found in the data can be retrieved from the graph, the graph is called *faithful* to the data.

The faithfulness property is often criticized, especially its validity. We will experimentally show in the following section that, even for 'nice' linear Gaussian models, the assumption indeed gets violated. Violation of faithfulness is regarded as a measure zero event (Pearl, 2000). This is, however, not true when working with real data. Due to limited sample sizes, accidental correlations occur and dependencies get 'weak' such that they are observed as independencies.

2 Violation of faithfulness

To illustrate how the faithfulness property gets violated, simulations were performed on linear Gaussian models. Experiments were performed on 50 randomly selected graphs with 50 nodes and 50 edges. For each such graph, a random structural equation model was constructed by selecting edge coefficients uniformly from $[0.5, 1.5] \cup [-1.5, -0.5]$. The standard deviation for the Gaussian disturbance term of the equations was selected uniformly from $[0.01, 1]$. The values of input nodes are randomly chosen from $[0, 1]$.

A random data set of 1000 cases was simulated for each of the models, to which the standard constraint-based PC algorithm was applied with depth 2 and significance level $\alpha = 0.05$ for the independence test based on partial correlation. The output graph was compared to the Markov equivalence pattern of the true DAG. The following table shows the average number of errors.

Adjacency false negatives	6.36
Adjacency false positives	7.7
Arrowhead false negatives	1.9
Arrowhead false positives	16.34

(Ramsey et al., 2006) showed that the correctness of the PC algorithm relies on adjacency and orientation faithfulness. We tested the validity of both assumptions on the simulated data. The following table shows that both assumptions do not hold. Orientation faithfulness is tested for all triples of nodes in which exactly 2 pairs are adjacent.

Violations of adjacency faithf.	6.36	50
Violations of orientation faithf.	43.86	92.9

3 Reasons of learning errors

In our attempt to find good explanations of the non-causal conditional independencies, we investigated the errors of the PC algorithm and (a) tried to identify a property of the model responsible for the error. Then we looked for (b) a rule, like the d-separation criterion, to infer the conditional independencies that are generated by the property. Next, we investigated if (c) the property can be detected in the data and if so, (d) whether the right causal relations can be identified.

The following properties were identified:

1. **Undetectable weak edge**: (a) is a direct causal relation between 2 variables for which the marginal dependency is not detected by the independence test under the given sample size. (b) All influences through the causal relation get undetected, as if the causal connection is not present. (c) It is undetectable since the adjacent variables do not become marginally dependent or when conditioned on another variable. It is a strong violation of adjacency faithfulness which cannot be solved.

2. **Detectable weak edge**: (a) the adjacent variables, say X and Y, are measured to be marginally independent, but they become dependent when conditioned on another variable, say Z. Moreover, (b) we demand, for consistency reasons, that all d-connected variables through $X - Y$ become dependent by conditioning on Z. It generates a violation of adjacency faithfulness and a violation of orientation-faithfulness, see also (Ramsey et al., 2006), (c) which is detectable, but (d) independencies alone do not always provide enough information to identify the right causal relations.

3. **Accidental correlation**: (a) a correlation between two variables is accidental, the variables are not causally connected in the model, but (b) are qualified as being dependent, even when conditioned on other variables. (c) An accidental correlation cannot be detected by a conditional independence test unless the significance level is increased.

4. **Information equivalent variables**: (a) two variables X and Y contain the same information about another variable Z (Lemeire, 2007). This happens with deterministic variables, but also with quasi-deterministic variables. (b) D-separation tells us which conditional independencies follow from a deterministic relation (Geiger, 1990). (c) It is easily detected by a violation of the intersection condition: $X \not\perp\!\!\!\perp Z$, $X \perp\!\!\!\perp Z \mid Y$ and $Y \perp\!\!\!\perp Z \mid X$ and (d) can be solved by comparing the complexity between the relations $X - Z$ and $Y - Z$ (Lemeire, 2007).

With this, the following errors could be related to one of the 4 above properties. The same simulations were made as in the previous section. The first column gives the number of errors, the other columns give the number of errors that can be explained by one of the 4 properties. The properties are identified based on the true graph.

	#	1	2	3	4
Adj. false neg.	6.4	0.28	0.5	0	2.9
Adj. false pos.	7.8	0	0	7.3	0
Arr. false neg.	2.1	0.06	0.1	0	0.9
Arr. false pos.	15	0	0	8.9	1.0

As shown in the table, accidental correlations and information equivalences are responsible for almost all false positive arrowheads. The (in)dependencies they generate lead to false v-structures.

4 Conclusions

Faithfulness is violated by other than causal properties of the system under limited sample size. We identified 4 properties that generate non-causal conditional independencies. The goal is now to detect these properties by patterns in the independencies and integrate them with Bayesian networks so that together they can explain the observed conditional independencies.

References

D. Geiger. 1990. *Graphoids: A Qualitative Framework for Probabilistic Inference*. Ph.D. thesis, University of California, Los Angeles.

Jan Lemeire. 2007. *Learning Causal Models of Multivariate Systems and the Value of it for the Performance Modeling of Computer Programs*. Ph.D. thesis, Vrije Universiteit Brussel.

Judea Pearl. 2000. *Causality. Models, Reasoning, and Inference*. Cambridge University Press.

Joseph Ramsey, Jiji Zhang, and Peter Spirtes. 2006. Adjacency-faithfulness and conservative causal inference. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, Arlington, Virginia. AUAI Press.