
Learning high-dimensional DAGs with latent and selection variables

Diego Colombo

Seminar for Statistics
ETH Zurich
8092 Zurich
Switzerland

Marloes H. Maathuis

Seminar for Statistics
ETH Zurich
8092 Zurich
Switzerland

Markus Kalisch

Seminar for Statistics
ETH Zurich
8092 Zurich
Switzerland

Thomas S. Richardson

Department of Statistics
University of Washington
Seattle, Washington 98195
USA

Introduction

We consider the problem of learning causal information between random variables in directed acyclic graphs (DAGs) when allowing arbitrarily many latent and selection variables. The Fast Causal Inference algorithm (FCI) (Spirtes *et al.*, 1999) has been explicitly designed to infer conditional independence and causal information in such settings. Despite its name, FCI is computationally very intensive for large graphs. Spirtes (2001) introduced a modified version of FCI, called Anytime FCI, which only performs conditional independence tests up to a pre-specified cut-off k . Anytime FCI is typically faster but less informative than FCI, but the causal interpretation of tails and arrowheads in its output is still sound. We propose an adaptation of Anytime FCI, called Adaptive Anytime FCI (AAFICI), where the cut-off k is set to the maximum size of the conditioning sets used to find the initial skeleton in FCI. Moreover, we propose a new algorithm, called *Really* Fast Causal Inference (RFCI), which has similar properties as AAFICI but is much faster for large sparse graphs. The complete paper is available at <http://arxiv.org/abs/1104.5617>.

The RFCI algorithm

The main difference between FCI and RFCI is that RFCI avoids conditional independence tests given subsets of “Possible-D-SEP”, which can become very large even for sparse graphs. Instead, RFCI performs some additional local tests before orienting v -structures and discriminating paths in order to ensure soundness. As a result, RFCI is much faster than FCI and AAFICI for sparse graphs. Although the output of RFCI can be slightly less informative than that of FCI, we prove that any causal information in the output of RFCI is still correct.

When are the algorithms identical?

We specify graphical conditions on an underlying DAG such that the outputs of FCI, AAFICI, and RFCI are

identical. Moreover, if the outputs are not identical, we infer properties of the edges that are present in the output of RFCI but not in that of FCI.

Consistency

We prove consistency of all algorithms in sparse high-dimensional settings. The sparsity conditions needed for consistency of RFCI and AAFICI are similar to the ones used in Kalisch and Bühlmann (2007) for consistency of the PC algorithm. The conditions for FCI are significantly stronger, due to the higher computational complexity of this algorithm.

Numerical Examples

For sparse graphs, the computational complexity of FCI is exponential, while AAFICI and RFCI are polynomial (worst case). To study the practical complexity of the algorithms, we compare them in a simulation study, where we also include several modifications of FCI and AAFICI that are identical to the original algorithms in the oracle versions but faster in the sample versions. We show that the number of estimation errors made by all algorithms are about the same. Moreover, our modifications of FCI and AAFICI shorten the computations considerably, but RFCI is the only feasible algorithm for large graphs.

References

- Kalisch, M. and Bühlmann, P. (2007) Estimating high-dimensional directed acyclic graphs with the PC algorithm. *J. Mach. Learn. Res.*, **8**, 613–636.
- Spirtes, P. (2001) An anytime algorithm for causal inference. In *Proc. of AISTATS 2001*, 213–221.
- Spirtes, P., Meek, C. and Richardson, T. S. (1999) An algorithm for causal inference in the presence of latent variables and selection bias. In *Computation, Causation and Discovery*, 211–252. MIT Press.