

## Learning Causal Structure from Data

Dr. Jan Lemeire (VUB) 03/11/2008.

In this ambitious field, we want to learn the underlying causal structure of a system under study. The causal structure is represented by a graph, in which the nodes represent the variables of interest and the links the direct causal relations. Graphical causal models are in fact Bayesian networks in which the links are interpreted as causal relations. The interpretation is based on the possibility to predict the effect of changes to the system by so-called interventions. While the theory of Bayesian networks (as representations of probability distributions) is generally accepted, causal structure learning is not. A major problem is that causal relations are not directly observable (which also gives philosophical problems in defining causation). Causal learning is based on the conditional independencies that follow from a causal structure. Basically, they follow from Markov chains and from v-structures. The learning algorithms can be divided in scoring-based and constraint-based algorithms. Personally, I find the latter the most attractive, since they also output the evidence by which the causal relations are found. Unfortunately, at present these algorithms are not yet robust since they rely on some unpractical assumptions.