Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2015:10**
http://www.math.su.se

Causal Inference through Structure Learning in Bayesian Networks

Simon Berggren*

June 2015

Abstract

This thesis is about structure learning in Bayesian Networks, and how this may be used for causal inference. A Bayesian Network is a graphical representation of a probability distribution, that provides a clear representation of conditional independences among the random variables. It consists of a graph and a set of probability tables. In the graph, each node represent a random variable, while edges and their orientations represent association and its nature. Causal inference is an analysis concerned with queries about cause and effect relationships. Specifically, in the phenomenon we wish to analyze, we use random variables to model the factors that describe this phenomenon, and infer causal associations among those variables.

Structure learning is a method for obtaining a Bayesian Network from data, i.e. given data consisting of a number of observations, where each observation consists of a realization of all the random variables in our model, the task is to infer a graph structure that represent the distribution of the variables. Hence, the main focus of this thesis will be to obtain a graph structure from data, such that the cause and effect relationships it exhibits, represents the corresponding causal associations in the underlying distribution. There are different approaches to structure learning in Bayesian Networks. In this thesis, focus is on the constraint-based approach, in which we use conditional independences, inferred from data, to build the graph structure. Since the number of required independence tests increases fast with the number of variables in our model, algorithms are necessary to handle the learning process.

One part of this thesis consists of a theoretical treatment of structure learning in Bayesian Networks, and in another part, I have composed and implemented my own version of a structure learning algorithm, SLBN, which is available for download at `https://github.com/SiboBerggren/SLBN`. Furthermore, simulation studies were performed to test the capabilities of structure learning, and of the SLBN algorithm. Altogether, we will see that structure learning is a powerful tool that may be used for causal inference, and that the SLBN algorithm successfully handles the learning process.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: simon.berggren@gmail.com. Supervisor: Michael Höhle.