Abstract:
Over the last years, vine copulas have gained more and more popularity in the statistical community as they allow for very flexible modelling of complex dependence structures. Using so-called pair copula constructions we can construct high-dimensional copulas by using bivariate building blocks. In this thesis we use the subclass of D-vine copulas in order to develop a new method for mean regression. We present five different methods for conditional mean estimation using the properties of D-vine copulas. These methods allow us to construct predictive models for a response variable given some observed covariates. Within this framework the copulas can either be estimated parametrically or nonparametrically. Moreover, we discuss a workaround for discrete data, since in this case the classical copula theory cannot be applied. Another currently hot topic are machine learning methods and in the context of mean regression especially gradient boosting is a very frequently used tool. However, often it is seen as a black-box and most of the literature lacks detailed mathematical explanations. Therefore, we will give an extensive exposition on gradient boosting for regression trees, including an introduction to regression trees and simple examples in order to clarify the methodology. Afterwards, we will show in an exhaustive simulation study that our newly developed D-vine copula based mean regression can achieve very good results in many different scenarios and can even outperform gradient boosting in some setups.
Finally, we will also apply the presented methods to a real-world data set.