YRD September 2017

Rebecca Marion

Title: Model Regularization for the Selection of Variable Groups in *Omics* Data

Abstract

Statistical classification models that predict patient disease states or subtypes based on "omics" data play an important role in "personalized medicine," making it possible to improve the quality of patient diagnosis and treatment. However, the complex structure of dependencies between variables in these "omics" data sets make it difficult to reliably identify the variables most predictive of disease. Correlated predictor variables tend to form groups that could represent a biological entity, such as a protein or metabolite, or a biological process. In such a case, it is important to select or exclude all variables in a given group so that the mechanisms of disease can be studied with greater precision and comprehensiveness. One popular approach in the literature for identifying important variables or variable groups is to impose constraints on the predictive model that induce *sparsity* (i.e. dependence on a reduced set of predictor variables). During this presentation, several model regularization methods of this type will be presented and compared. The empirical performance of these methods, as demonstrated in several simulation studies, will also be highlighted.

Stefka Asenova (Supervisor – Johan Segers)

Title : Graphical Models and Extremes

Abstract:

Graphical models form a class of statistical models designed for analyzing ensembles of stochastic variables whose joint law is determined by a set of conditional independence relations. A graph is an Independency map of the joint probability law if graphical separation implies conditional independence. Such structures, commonly known in literature, are Markov fields and Bayesian networks. When interest is in extreme values of the variables the analysis must include both the theory of Multivariate Extremes and Graphical models.

Extreme value theory is the branch of probability theory and statistics which aims at providing models for rare events and for observations that occur with low frequency but have a potential high impact, i.e., for the tails of statistical distributions. Examples include high water levels leading to flooding, financial market turmoil, and catastrophe insurance claims. Of noteworthy importance is the issue of tail dependence, that is, the propensity of such extreme values to occur simultaneously in many variables at once.

The aim of the project is then to develop models for extreme values of random variables whose dependence structure can be represented by a graph via conditional independence relations.

The current research focuses on tree structures. The goal is to combine the theoretical results about asymptotic convergence of a tree graphical model and one of the novel estimators of multivariate tail dependence to obtain an estimator of extreme dependencies in a regularly varying tree model. The scope of the paper is to illustrate how both theories can be combined to draw inference on the tail dependence in case a tree structure is known or assumed.

Dimitra Kiryakopoulu

The title of my talk is “Exponential-type GARCH models with linear-in-variance risk premium” (joint paper with Prof. Hafner). It is both theoretical and applied J. However, one of the main contributions is the theoretical part, that is accompanied by an empirical section to be complete.

Abstract:

One of the implications of the intertemporal capital asset pricing model (CAPM) is that the risk premium of the market portfolio is a linear function of its variance. Yet, estimation theory of classical GARCH-in-mean models with linear-in-variance risk premium requires strong assumptions and is incomplete. We show that exponential-type GARCH models such as EGARCH or Log-GARCH are more natural in dealing with linear-in-variance risk premia. For the popular and more difficult case of EGARCH-in-mean, we derive conditions for the existence of a unique stationary and ergodic solution and invertibility following a stochastic recurrence equation approach. We then show consistency and asymptotic normality of the quasi maximum likelihood estimator under weak moment assumptions. An empirical application estimates the dynamic risk premia of a variety of stock indices using both EGARCH-M and Log-GARCH-M models.

Kassu Mehari Beyene, Anouar El Ghouch and Abderrahim Oulhaj

Time-dependent ROC Curve Estimation with Cure Fraction

Abstract

The ROC curve and its summary measure AUC are the two com-monly used tools to evaluate the classification accuracy of a continuous variable for a binary outcome. The time-dependent ROC curves have been used to assess the predictive ability of diagnostic markers for survival analysis. Several authors proposed methods to estimate the time-dependent ROC curves and AUC for a survival analysis. The va-lidity of the estimators from these methods relay on some assumptions. One of the assumptions is that, all subjects of the study population is susceptible to the event of interest and will eventually experience this event if the follow-up period is sufficiently long. However, this assump-tion may not be valid in many cases and hence studying the sensitivity of the estimators for the violation of this assumption is of substantial interest. The main aim of this article is to assess the validity of the time-dependent ROC curve and its summary measure AUC for data with cured subjects. An in depth simulations was conducted to study the performance of the estimator. The simulation studies make evi-dent that, when the marker is known or correctly estimated, the simple method of Li et al. (2016) is insensitive to the violation of the above assumption and therefore result in valid estimates for the classification accuracy measures.