Plenary Conferences

Modeling informative interval-censoring mechanism in regression models with Student's-t distribution

29 Nov
9:00

Victor Hugo Lachos
University of Connecticut

In many studies, censored data are collected, i.e., the exact true value is recorded only if it falls within an interval range, so the responses can be either left, interval, or right censored. Censored linear regression models are routinely used to analyze these types of data. However, most of these models assume that the censoring is non-informative (or random). In this work, we propose an EM-type algorithm for computing the maximum likelihood estimates for the Interval-censored linear regression models with informative (or not at random) censoring. Our algorithm uses analytical expressions at the E-step, which rely on formulas for the mean and variance of a truncated Student's-t distribution. The standard errors, the prediction of unobserved values of the response, and the log-likelihood function are obtained as a by-product. The proposed methodology is illustrated through the analyses of simulated and real data applications. conference

---

Cutoff-aware BART for Estimating Heterogeneous Treatment Effects in Regression Discontinuity Designs

30 Nov
9:00

Hedibert Freitas Lopes
INSPER

This paper proposes a modification of the Bayesian Causal Forest algorithm (Hahn et al., 2020) - itself an extension of the BART algorithm (Chipman et al., 2010) - which uses a novel regression tree prior that incorporates the unique structure of regression discontinuity designs. Specifically, we add constraints to the tree splitting process that assure overlap within a narrow band surrounding the running variable cutoff value (where the treatment effect is identified). We show that unmodified BART and BCF models estimate RDD treatment effects poorly, while our modified model accurately recovers treatment effects at the cutoff. At the same time, our modified model retains the inherent flexibility of all BART-based models, allowing it to effectively explore heterogeneous treatment effects. Simulation studies indicate that the new approach improves upon traditional local polynomial regression on both simple and complex data generating processes in terms of estimation

error, coverage, and interval length for both average and conditional average treatment effects. We illustrate the new method by analyzing data studied originally by Lindo et al. (2010) to estimate the effect of academic probation on university students' GPA; we find an average increase of 0.15 in GPA for students whose previous semester GPA lied just below the probation cutoff. This is joint work with Rafael Alcantara, Meijia Wang and P. Richard Hahn.

---

01 Dec
9:00

A penalization method to estimate the intrinsic dimensionality of data

Daniela Rodriguéz
Universidad de Buenos Aires

In this presentation, I will introduce a novel method based on penalization to estimate the intrinsic dimensionality of data. Our approach assumes that the data follow an extension of the Probabilistic Principal Components Model beyond the Gaussian case. I will delve into the theoretical results of our proposal and also demonstrate the competitive advantages of our methodology compared to several existing approaches in the literature.

Talks

The Role of the Marshall-Olkin Model in Network Reliability

29 Nov
10:30

Javiera Barrera
Universidad Adolfo Ibáñez

The Marshall-Olkin (MO) random vector is a tuple that takes positive values such that each marginal variable distribution is exponential and preserves the memory loss property. Proposed in 1967, it has captured the interest in several areas, given its capacity to capture strong and weak dependence. The model is very flexible and has many subfamilies with good properties that make it appealing from a technical and modeling point of view. In reliability, the MO model is used to represent components' lifetime. The model allows us to express simultaneous failures while maintaining the tractability to simulate and calibrate parameters. In this talk, we will discuss the mathematical structure of the model, the subfamilies, and some applications to analyze the network reliability.

Recovery Communities in Weighted Networks

29 Nov
11:00

Andressa Cerqueira
UFSCar

Network models have received increasing attention from the statistical community, in particular in the context of analyzing and describing the interactions of complex random systems. In this context, community structures can be observed in many networks where the nodes are clustered in groups with the same connection patterns. In this talk, we address the community detection problem for weighted networks in the case where, conditionally on the node labels, the edge weights are drawn independently from a Gaussian random variable with mean and variance depending on the community labels of the edge endpoints. We will present a fast and tractable EM algorithm to recover the community labels that achieves the optimal error rate.

---

01 Dec 10:30

Adjusted profile likelihoods for the Log-Birnbaum-Saunders regression model shape parameter under type II censoring

Audrey Cysneiros
UFPE

In this paper we obtain adjusted profile maximum likelihood estimators for the Log-Birnbaum-Saunders regression model shape parameter under type II data censoring. We consider the adjustments to the profile likelihood function proposed by Barndorff-Nielsen (1983), Cox and Reid (1987), Fraser and Reid (1995), Fraser et al. (1999) and Severini (1998, 1999). Several bootstrap confidence intervals are considered and also we obtain modified profile likelihood ratio tests. Monte Carlo simulation results on point estimation, interval estimation and hypothesis testing inference are reported. The results showed that both the tests and the estimators based on modified versions of the profile likelihood developed had superior performance in small samples in comparison with the estimators based on the unmodified version. Finally, an application to real data is presented.

---

01 Dec 11:00

Limit theorems for the simple parking process

Sandro Gallo
UFSCar

We consider the parking process on Z^2 with a simple occupancy scheme, which is defined as follows. Initially, all sites in $L(n) := -n, \ldots, n^2$ are empty. A site is chosen at random in L(n) and if all its nearest neighbor sites are empty, the site is occupied. Once occupied, the site remains so forever. The process continues until all sites in L(n) are either occupied or have at least one of their nearest neighbors occupied. The final configuration (occupancy) of L(n) is called the jamming limit and is denoted by P(n). When n tends to infinity, P(n) tends to P, the final configuration on Z^2, which is called the thermodynamic jamming limit. In this talk I will discuss statistical properties of the thermodynamic jamming limit, such as law of large numbers and central limit theorem for the number of occupied sites in the n-box.

Short-course

An Introduction to GAMLSS

30 Nov
13:30

Mikis Stasinopoulos

The Generalized Additive Models for Location, Scale and Shape (GAMLSS) introduced almost 20 year ago by Rigby and Stasinopoulos (2005), are a very general framework for univariate regression. Their novelty arise from the fact that all the parameters of the assumed parametric distribution of the response (target) can be modelled as functions of the explanatory variables (features). This allows modelling a response variable with high skewness or kurtosis. The assumed distribution can be any theoretical distribution as described by Rigby et al. (2019). There are different ways for modelling all the distribution parameters using explanatory variables. These include linear terms, smoothing terms and any sensible Machine Learning technique like neural networks, LASSO, principal component regression etc., see Stasinopoulos et al. (2017). There are also different ways of fitting the model which includes the classical (using penalized likelihood), the Bayesian using MCMC, or boosting, see Stasinopoulos et al. (2023).

This short course is designed for practitioners and applied statisticians which would like to know how to go about modelling their data set using GAMLSS. The short course will discuss exploratory checking of the data and problems associated with the creation of a suitable data set for the analysis, to the selection of the right distribution, and the selection of the right explanatory variables modelling each parameter. Finally the short course will describe how to interpret the fitted model(s) and how it can be used for predictions. More information about GAMLSS can be found at `http://www.gamlss.com/`.

References

Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized Additive Models for Location, Scale and Shape (with discussion). *Appl. Statist.*, **54**, 507-554.

Stasinopoulos D. M., Rigby R. A., Heller G., Voudouris V. and De Bastiani. F. (2017) *Flexible Regression and Smoothing: Using GAMLSS in R.* Chapman and Hall/CRC.

Rigby R. A., Stasinopoulos D. M., Heller G., and De Bastiani. F. (2019). *Distributions for Modelling Location, Scale and Shape: Using GAMLSS in R.* Chapman and Hall/CRC.

Stasinopoulos D. M., Kneib T., , Klein N., Meir A., and Heller G. *Generalized Additive Models for Location Scale and Shape: A Distributional Regression Approach with Applications.* To appear