Resumos

Worskshop on Time Series, Wavelets and High Dimensional Data

STODAD2019

Campinas - SP

August 29-30, 2019

Projeto Temático FAPESP 2018/04654-9
Séries Temporais, Ondaletas e Dados de Alta Dimensão

Abstract
Invited Speakers

Jianqing Fan, Princeton University, USA

Title: Noisy matrix completion: Understanding statistical errors of convex relaxation via Nononconvex optimization

Abstract: This paper studies noisy low-rank matrix completion in presence of noise. When the rank of the unknown matrix is a constant, we demonstrate that the convex programming approach achieves near-optimal estimation errors — in terms of the Euclidean loss, the entrywise loss, and the spectral norm loss — for a wide range of noise levels. All of this is enabled by bridging convex relaxation with the nonconvex Burer–Monteiro approach, a seemingly distinct algorithmic paradigm that is provably robust against noise. More specifically, we show that an approximate critical point of the nonconvex formulation serves as an extremely tight approximation of the convex solution, allowing us to transfer the desired statistical guarantees of the nonconvex approach to its convex counterpart.

(Based on the joint work with Yuejie Chi, Yuxin Chen, Cong Ma, Yulin Yang)

Qiwei Yao, London School of Economics

Title: Estimation of Subgraph Densities in Noisy Networks

Abstract: While it is common practice in applied network analysis to report various standard network summary statistics, these numbers are rarely accompanied by some quantification of uncertainty. Yet any error inherent in the measurements underlying the construction of the network, or in the network construction procedure itself, necessarily must propagate to any summary statistics reported. Here we study the problem of estimating the density of an arbitrary subgraph, given a noisy version of some underlying network as data. Under a simple model of network error, we show that consistent estimation of such densities is impossible when the rates of error are unknown and only a single network is observed. Next, focusing first on the problem of estimating the density of edges from noisy networks, as a canonical prototype of the more general problem, we develop method-of-moment estimators of network edge density and error rates for the case where a minimal number of network replicates are available. These estimators are shown to be asymptotically normal as the number of vertices increases to infinity. We also provide confidence intervals for quantifying the uncertainty in these estimates based on either the asymptotic normality or a bootstrap scheme. We then present a generalization of these results to higher-order subgraph densities, and illustrate with the case of two-star and triangle densities. Bootstrap confidence intervals for those high-order densities are constructed based on a new algorithm for generating a graph with pre-determined counts for edges, two-stars, and triangles. The algorithm is based on the idea of edge-rewiring, and is of some independent interest. We illustrate the use of the proposed methods in the context of gene coexpression networks.

Robert Lund, Clemson University

Title: Multiple Breakpoint Detection: Mixing Documented and Undocumented Changepoints

Abstract: This talk presents methods to estimate the number of changepoint time(s) and their locations in time series when prior information is known about some of the changepoint times. A Bayesian version of a penalized likelihood objective function is developed from minimum description length (MDL) information theory principles. Optimizing the objective function yields estimates of the changepoint number(s) and location time(s). Our MDL penalty depends on where the changepoint(s) lie, but not solely on the total number of changepoints (such as classical AIC and BIC penalties). Specifically, configurations with changepoints that occur relatively closely to one and other are penalized more heavily than sparsely arranged changepoints. The techniques allow for autocorrelation in the observations and mean shifts at each changepoint time. This scenario arises in climate time series where a "metadata" record exists documenting some, but not necessarily all, of station move times and instrumentation changes. Applications to climate time series are presented throughout, including some recent controversies about Atlantic hurricane changes.

Junyuan Chang, FUSE, China

Title: A new scope of penalized empirical likelihood with high-dimensional estimating equations

Abstract: Statistical methods with empirical likelihood (EL) are appealing and effective especially in conjunction with estimating equations for flexibly and adaptively incorporating data information. It is known that EL approaches encounter difficulties when dealing with high-dimensional problems. To overcome the challenges, we begin our study with investigating high-dimensional EL from a new scope targeting at high-dimensional sparse model parameters. We show that the new scope provides an opportunity for relaxing the stringent requirement on the dimensionality of the model parameters. Motivated by the new scope, we then propose a new penalized EL by applying two penalty functions respectively regularizing the model parameters and the associated Lagrange multiplier in the optimizations of EL. By penalizing the Lagrange multiplier to encourage its sparsity, a drastic dimension reduction in the number of estimating equations can be achieved. Most attractively, such a reduction in dimensionality of estimating equations can be viewed as a selection among those high-dimensional estimating equations, resulting in a highly parsimonious and effective device for estimating high-dimensional sparse model parameters. Allowing both the dimensionalities of model parameters and estimating equations growing exponentially with the sample size, our theory demonstrates that our new penalized EL estimator is sparse and consistent with asymptotically normally distributed nonzero components. Numerical simulations and a real data analysis show that the proposed penalized EL works promisingly.

Marcelo Fernandes, FGV-SP

Title: Nonparametric testing of conditional independence using asymmetric kernels

Abstract: Statistical tools for testing conditional independence between X and Y given Z are developed. In particular, we test whether the conditional density of X given Y and Z is equal to the conditional density of X given Z only. We gauge the closeness between these conditional densities using a generalized entropic measure. To avoid degeneracy issues, we transform the variables of interest (X,Y,Z) to bound them in the unit interval, and then estimate their conditional densities using beta kernels. The latter are convenient because they are free of boundary bias. We show that our test statistics are asymptotically normal under the null hypothesis as well as under local alternatives. We assess the finite-sample properties of our entropic-based tests of conditional independence through Monte Carlo simulations.

Chang Chian, USP

Title: Estimating the trace-variogram in the ordinary kriging method for functional data using Legendre-Gauss quadrature

Abstract: In several applications, as ecological and environmental data analysis, information is spatially indexed curves which are denominated Spatial Functional Data. The main goal of this paper is to supply a simpler approach to interpolate curves. More precisely, we propose the use of Legendre-Gauss quadrature to estimate the trace variogram in the ordinary kriging method introduced by Giraldo (2011). Additionally, we also propose a methodology to construct confidence interval for the curve. We use simulated datasets to compare the proposed estimation procedure with the usual method of estimation. As results, a similar bias and a lower mean square error are observed to support the proposed estimation process. The novel estimation methodology is applied to dataset on dailly mean temperature from 35 weather stations of Canadian Maritime Provinces.

(joint work with Gilberto P. Sassi)

Luiz K. Hotta, Unicamp

Title: Forecasting Conditional Covariance Matrices in High-Dimensional Time Series: a General Dynamic Factor Approach

Abstract: Based on a General Dynamic Factor Model with in finite-dimensional factor space, we develop new estimation and forecasting procedures for conditional variance matrices in high-dimensional time series. The performance of our approach is evaluated via Monte Carlo experiments, yielding excellent finite-sample properties. The new procedure is used to construct minimum variance portfolios in from a high-dimensional panel of assets. The results are shown to achieve better out-of-sample portfolio performance than alternative existing procedures.

(joint work with Carlos Trucíos, Mauricio Zevallos, Pedro L. Valls Pereira and Marc Hallin)

Marcelo Medeiros, PUC-RJ

Title: BooST: Boosting Smooth Transition Regression Trees for Partial Effect Estimation in Nonlinear Regressions

Abstract: In this paper, we introduce a new machine learning (ML) model for nonlinear regression called the Boosted Smooth Transition Regression Trees (BooST), which is a combination of boosting algorithms with smooth transition regression trees. The main advantage of the BooST model is the estimation of the derivatives (partial effects) of very general nonlinear models. Therefore, the model can provide more interpretation about the mapping between the covariates and the dependent variable than other tree-based models, such as Random Forests. We show the consistency of the partial derivative estimates, and we present examples with both simulated and real data.

Flavio Ziegelmann, UFRGS

Title: Nonparametric Frontier Estimation

Abstract: In this paper we analyze a nonparametric model for estimating production frontiers with multi-dimensional inputs, with some advantages over traditional frontier estimators as Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH). In comparison to Martins-Filho and Yao (2007) approach, our method is simpler, more general and requires less restrictive assumptions. We also show consistency and asymptotic distributional theory of our proposed estimator under standard assumptions in the multi-dimensional input setting. The finite-sample performance of our proposed estimator is illustrated in a simulation study where we compare it with Martins-Filho and Yao's, DEA and FDH.

Pedro A. Morettin, USP

Title: Wavelet Estimation of Copulas for Time Series

Abstract: In this paper, we consider estimating copulas for time series, under mixing conditions, using wavelet expansions. The proposed estimators are based on estimators of densities and distribution functions. Some statistical properties of the estimators are derived and their performance assessed via simulations. Empirical applications to real data are also given.

KEYWORDS: copula, density, time series, wavelet, wavelet estimators

Mauricio Zevallos, Unicamp

Title: Estimation of ARFIMA models: a minimum distance approach

Abstract: The paper proposes a new minimum distance estimator (MDE) for Gaussian ARFIMA processes with long-memory parameter in the interval (-1/2,1/2). The MDE method is based on the minimization of the distance between sample and population autocovariance differences. This permits, in addition, the simultaneous estimation of the variance of the errors. It is shown that the new estimator satisfies a central limit theorem and Monte Carlo experiments indicate that the proposed estimator performs very well. The proposed method is illustrated with the estimation of real-life time series.

Keywords: Autocovariances, fractional noise.