

Internal mini-workshop

Advances in Big-Data Modelling

March, 24, 2023, 9.30am -1.00pm

Meeting Room 1 - San Giobbe Economics Campus

9.30- 10.10

Radu V. Craiu (Department of Statistical Sciences, University of Toronto) *Copula modelling of serially correlated multivariate data with hidden structures*

I will consider a copula-based extension of the hidden Markov model. At each measuring time, a vector of observations is recorded for each unit in the sample. The joint model induced by the copula extension allows decoding of the hidden states based on information from multiple observations but brings forth computational challenges. These are tackled using a theoretically justified variation of the EM algorithm developed within the framework of inference functions for margins. The method is illustrated using numerical experiments and a real example. This is joint work with Dr. Vianey Leos-Barajas (Toronto) and Robert Zimmerman (Toronto).

10.10-10.50

Ovielt Baltodano (Department of Economics, Ca' Foscari University of Venice), *Dynamic Stochastic Block Models*

10.50-11.20

Coffee Break

11.20-12.00

Lei Sun (Department of Statistical Sciences, University of Toronto), *Back to the future: 'simple' linear regression to generalize Kruskal-Wallis, Levene's, and Pearson's Hardy-Weinberg equilibrium tests.*

Regression remains an important framework in the era of big and complex data. In this talk, I will present some recent examples where we resort to the simple linear regression and its celebrated extensions in novel settings, including generalizing Kruskal-Wallis, Levene's, and Pearson's Chisq Hardy-Weinberg equilibrium tests. The Eureka moment

came while reading Wu and Guan's (2015) comments on our generalized Kruskal-Wallis test (Acar and Sun 2013, Biometrics). Wu and Guan presented an alternative "rank linear regression model and derived the proposed GKW statistic as a score test" Indeed, the regression framework eases the derivation and facilitates further extensions. More recently, we turned our attention to extending Levene's variance test for dependent sample with group uncertainty; this test is useful for latent Gene-Environment interaction analysis where data on E is unavailable. While a direct modification of the original test statistic is challenging, I will demonstrate how a two-stage regression framework eases the extension, leading to a generalized joint location-scale test (Soave and Sun 2017, Biometrics). Finally, I will discuss how to formulate the classical Pearson's Chisq test of Hardy-Weinberg equilibrium as a linear regression, and how this formulation can be utilized to generalize the locally most powerful allelic association test (Zhang and Sun 2022, Biometrics), develop methods for the complex X chromosome that is often omitted in the current genome-wide association studies, and beyond. In summary, the crux of each work is reformulating the problem as a regression!

12.00-12.40

Qing Wang (Department of Economics, Ca' Foscari University of Venice), *A Bayesian Tensor regression model*

13.00

Lunch

Organizing committee:

Ovielt Baltodano, Roberto Casarin, Davide Raggi

San Giobbe, Cannaregio 873, 30121 – Venezia | Tel: +39 041 234 9254
Marianna Morelli, e-mail: centro.vera@unive.it