Rencontre estivale du Laboratoire de Statistique
Summer Meeting of the Statistics Laboratory

Debbie Dupuis (HEC Montréal)

Measuring Correlation in the Presence of Spikes

The cost of electricity varies across the zones of the New York State electric system. While fair and open access to the electrical grid is sought, we show that that residents currently do not equally benefit, or suffer, from price changes. Upcoming major investments in the grid offer an opportunity to rectify these inequalities, but only if we understand the price-change propagation dynamics for the current underlying infrastructure. We study these dynamics, estimating the partial correlations between changes in electricity prices in connected zones. We develop and investigate a robust exponentially weighted correlation estimator that performs well in the presence of electricity price spikes and can track a rapidly-changing time-varying correlation. We show that price-change partial correlations are mostly positive, but can also be negative, and provide new insight into price-change dynamics within the grid that cannot be extracted from the price-setting algorithm or obtained from available transmission capability data.

Russell Steele (McGill)

A statistician's view of singular learning: model selection for latent variables models

In this talk, I will discuss model selection approaches for latent variable models using tools from the field of singular learning developed by Sumio Watanabe. The singular learning approach combines tools from algebraic geometry, complex analysis and empirical process theory in order to better characterize the large sample behaviour of existing and to propose novel model selection criteria. My talk will contain two primary parts. In the first part, I will give an overview of the basic ideas behind singular learning and emphasize key results from the literature. In the second part, I will demonstrate how singular learning methods can be applied to two of the most commonly used latent variable models (hierarchical linear models and finite mixture models) and compare results to what is obtained via existing model selection criteria.

Aurélie Labbe (McGill)

Principal component of explained variance: an efficient and optimal data dimension reduction framework for association studies

The genomics era has led to an increase in the dimensionality of the data collected to investigate biological questions. In this context, dimension-reduction techniques can be used to summarize high-dimensional signals into low-dimensional ones, to further test for association with one or more covariates of interest. This talk revisits one such approach, previously known as Principal Component of Heritability and renamed here as Principal Component of Explained Variance (PCEV). As its name suggests, the PCEV seeks a linear combination of outcomes in an optimal manner, by maximising the proportion of variance explained by one or several covariates of interest. By construction, this method optimises power but limited by its computational complexity, it has unfortunately received little attention in the past. Here, we propose a general analytical PCEV framework that builds on the assets of the original method, i.e. conceptually simple and free of tuning parameters. Moreover, our framework extends the range of applications of the original procedure by providing a computationally simple strategy for high-dimensional outcomes, along with exact and asymptotic testing procedures that drastically reduce its computational cost. We investigate the merits of the PCEV using an extensive set of simulations. Furthermore, the use of the PCEV approach will be illustrated using three examples taken from the epigenetics and brain imaging areas.

David Haziza (Université de Montréal)

Multiply robust imputation procedures for the treatment of item nonresponse in surveys

Item nonresponse in surveys is often treated through some form of imputation. In this paper, we introduce the concept of multiply robust imputation procedures in the context of finite population sampling, which is closely related to the concept of multiple robustness proposed by Han and Wang (2013). Multiple robustness can be viewed as an extension of the concept of double robustness. In practice, multiple nonresponse models and multiple imputation models may be fitted, each involving different subsets of covariates

and possibly different link functions. An imputation procedure is said to be multiply robust if the resulting estimator is consistent if all but one model are misspecified. Multiply robust point and variance estimators are proposed in the case of non-negligible sampling fractions. Extension to random and fractional imputations are discussed. Finally, we present the results of a simulation study assessing the performance of the proposed point and variance estimators.

Joint work with Sixia Chen, University of Oklahoma.

Lajmi Lakhal-Chaieb (Université Laval)

Assessment of familial genetic cancer risk in the presence of competing risks

In this presentation, an association model to estimate the penetrance (risk) of multiple cancers in the presence of competing risks is proposed. The association between multiple events is modelled via a copula and a proportional hazards model is specified for each competing event. This work is motivated by the analysis of successive cancers for people with Lynch Syndrome and the analysis of breast and ovarian cancers for women with a mutation in BRCA1/2 genes. The proposed inference procedure is adapted to handle missing genetic covariates and selection bias, induced by the data collection protocol of the data at hand. The performance of the proposed estimation procedure is evaluated by simulations and its use is illustrated with data sets from the Colon Cancer Family Registry (Colon CFR) and the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA).