

Rencontre automnale du laboratoire de statistique 2021

29 octobre 2021

Fall Meeting of Statlab 2021

October 29, 2021

Horaire / Schedule

12 :30-13 :30 Réunion d'affaires Statlab (réservée aux membres) / *Statlab Business Meeting (Members only)*

13 :30 - 14 :00 Asad Haris (McGill University), *conference in English*

Title: A Targeted Approach to Confounder Selection for High-Dimensional data.

Abstract: We consider the problem of selecting confounders for adjustment from a potentially large set of covariates, when estimating a causal effect. Recently, the high-dimensional Propensity Score (hdPS) method was developed for this task; hdPS ranks potential confounders by estimating an importance score for each variable and selects the top few variables. However, this ranking procedure is limited: it requires all variables to be binary. We propose an extension of the hdPS to general types of response and confounder variables. We further develop a group importance score, allowing us to rank groups of potential confounders. The main challenge is that our parameter requires either the propensity score or response model; both vulnerable to model misspecification. We propose a targeted maximum likelihood estimator (TMLE) which allows us to use nonparametric, machine learning tools for fitting these intermediate models. We establish asymptotic normality of our estimator, which consequently allows constructing confidence intervals. We complement our work with numerical studies on simulated and real data.

14 :00 - 14 :30 Bouchra Nasri (École de santé publique de l'Université de Montréal), *présentation en français*

Title: Test of serial dependence for multivariate time series with arbitrary distributions

Abstract: In this talk, I will present new tests of serial independence for a fixed number of consecutive observations from a stationary time series, first in the univariate case, and then in the multivariate case. The common distribution function is not assumed to be continuous, and the tests statistics are based on the multilinear copula process. Finite sample experiments will be presented to show the power of the proposed tests as well as a case study to illustrate the usefulness of the proposed methodologies.

14 :30 - 15 :00 Mamadou Yauck (McGill University), *conference in English*

Title: Recent developments in respondent-driven sampling

Abstract: Respondent-driven sampling (RDS) is a variant of link-tracing, a sampling technique for surveying hard-to-reach communities that takes advantage of community members' social networks to reach potential participants. While the RDS sampling mechanism and associated methods of adjusting for the sampling at the analysis stage are well-documented in the statistical sciences literature, methodological focus has largely been restricted to estimation of population means and proportions (e.g. prevalence). In this talk, I will present some recent methodological developments in the study of associational questions, the estimation of uncertainty and the assessment of the validity of inference when the social network is partially observed.

15:30 – 16:30 Colloque des sciences mathématiques du Québec

Tiffany Timbers (University of British Columbia), *conference in English*

Title: Opinionated practices for teaching reproducibility: motivation, guided instruction and practice

Abstract: In the data science courses at the University of British Columbia, we define data science as the study, development and practice of reproducible and auditable processes to obtain insight from data. While reproducibility is core to our definition, most data science learners enter the field with other aspects of data science in mind, for example predictive modelling, which is often one of the most interesting topic to novices. This fact, along with the highly technical nature of the industry standard reproducibility tools currently employed in data science, present out-of-the gate challenges in teaching reproducibility in the data science classroom. Put simply, students are not as intrinsically motivated to learn this topic, and it is not an easy one for them to learn. What can a data science educator do? Over several iterations of teaching courses focused on reproducible data science tools and workflows, we have found that providing extra motivation, guided instruction and lots of practice are key to effectively teaching this challenging, yet important subject. Here we present examples of how we deeply motivate, effectively guide and provide ample practice opportunities to data science students to effectively engage them in learning about this topic.