

Rencontre estivale du Laboratoire de statistique / *Spring Meeting of Statlab*
Jeudi le 20 juin 2019 / Thursday, June 20, 2019
Université de Sherbrooke

Horaire / Program

- . 9h à 9h50 : Benjamin Charlier, Professeur, IMAG, Université de Montpellier

Kernel Operations on the GPU, with autodiff, without memory overflows.

The KeOps library lets you compute generic reductions of large 2d arrays whose entries are given by a mathematical formula. It combines a tiled reduction scheme with an automatic differentiation engine, and can be used through Matlab, NumPy or PyTorch backends. It is perfectly suited to the computation of Kernel dot products and the associated gradients, even when the full kernel matrix does *not* fit into the GPU memory.

- . 9h50 à 10h20 : Pause-café

- . 10h20 à 11h10: Mireille Schnitzer, Professeure, Université de Montréal

Methodological considerations for the analysis of relative treatment effects in multi-drug-resistant tuberculosis from fused observational studies.

Multi-drug-resistant tuberculosis (MDR-TB) is defined as strains of tuberculosis that do not respond to at least the two most used anti-TB drugs. After diagnosis, the intensive treatment phase for MDR-TB involves taking several alternative antibiotics concurrently. The Collaborative Group for Meta-analysis of Individual Patient Data in MDR-TB has assembled a large, fused dataset of over 30 observational studies comparing the effectiveness of 15 antibiotics. The particular challenges that we have considered in the analysis of this dataset are the large number of potential drug regimens, the resistance of MDR-TB strains to specific antibiotics, and the identifiability of a generalized parameter of interest though most drugs were not observed in all studies. In this talk, I describe causal inference theory and methodology that we have applied or developed for the estimation of treatment importance and relative effectiveness of different antibiotic regimens with a particular emphasis on targeted learning approaches.

- . 11h10 à 12h00 : Khader Khadraoui, Professeur, Université Laval

Un seuillage adaptatif optimal pour la connectivité et la détection de communautés dans un réseau social aléatoire géométrique.

Un problème commun à de nombreuses disciplines est celui d'étudier correctement les réseaux complexes. La recherche sur ce problème apparaît en mathématiques appliquées (systèmes dynamiques gouvernés par des EDOs et EDPs), probabilités et statistique (modèles de graphes aléatoires) et en ingénierie informatique (apprentissage statistique). Un outil moins familier dans ce domaine est le modèle individu centré ("Individual Based Model" ou "IBM"). La théorie de l'IBM est un domaine interdisciplinaire en pleine croissance avec un très large spectre de motivations et d'applications. En considération des réseaux complexes, nous avons récemment caractérisé les systèmes de réseaux sociaux par l'individualité des membres et des mécanismes d'interaction localisés entre les membres grâce à des noyaux de dispersion spatialisés ce qui a conduit globalement à un comportement régulier du système. Durant cette présentation, nous prouvons que le nombre de membres isolés dans le réseau suit asymptotiquement une loi de Poisson en développant une méthode de discrétisation du carré unité et une inégalité de concentration. Nous établissons le seuil

optimal associé à cette approximation et nous montrons que ce résultat est valable pour la classe des fonctions de connectivité qui s'annulent au-delà du seuil. De plus, nous tirons parti de la distribution asymptotique des membres isolés pour construire une méthode de détection de communautés (classification) dynamique et adaptative adaptée à notre réseau.

. 12h00 à 13h40 : Lunch

. 13h40 à 14h30 : Geneviève Lefebvre, Professeure, UQÀM

Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: moving towards exact regression-based approaches

In the binary outcome framework to causal mediation, standard expressions for the natural direct and indirect effect odds ratios (ORs) are established from a logistic model by invoking several approximations that hold under the rare-disease assumption. Such ORs are expected to be close to corresponding effects on the risk ratio (RR) scale based on a log-binomial model, but the robustness of interpretation to this assumption has not been investigated. The objective was to report on mediation results from these two models when the (marginal) incidence of the outcome is $<10\%$. Standard (approximate) ORs and RRs were estimated using data on pregnant asthmatic women from Québec. Prematurity and low-birthweight were the mediator and outcome variables, respectively, and two binary exposure variables were considered: treatment to inhaled corticosteroids and placental abruption. Exact ORs were also derived and estimated using a contributed SAS macro. Simulations which mimicked our data were subsequently performed to replicate findings.

Many approximate ORs and RRs estimated from our data did not closely agree. Approximate ORs were systematically observed farther from RRs in comparison with exact ORs. Exact OR estimates were very close to RR estimates for exposure to inhaled corticosteroids, but less so for placental abruption. Approximate OR estimators also exhibited important bias and undercoverage in simulated scenarios which featured a strong mediator-outcome relationship. These results pave the way for exact estimators that do not rely on the rare-disease assumption.

Joint work with Mariia Samoilenko (UQAM). The presentation will be given in French with English slides.

. 14h30 à 15h20 : Alejandro Murua, Professeur, Université de Montréal

Un modèle bayésien semi-paramétrique pour le regroupement des bi-grappes

Motivés par les classes de problèmes fréquemment rencontrés dans l'analyse des données d'expression génique, nous proposons un modèle bayésien semi-paramétrique permettant de détecter les bi-grappes (biclusters), c'est-à-dire des sous-ensembles d'individus partageant des patrons similaires sur un ensemble de conditions. Notre approche est basée sur le modèle des carrés écossais proposé par Lazeroni et Owen (2002). En supposant une loi a priori de type bâton tronqué (truncated stick-breaking prior), nous estimons également le nombre de bi-grappes dans les données dans le cadre de l'inférence. Les résultats d'une étude de simulation montrent que le modèle est capable de détecter correctement les bi-grappes, et que sa performance est très bonne par rapport à certaines approches concurrentes. La flexibilité de la loi a priori proposée est démontrée par des applications à l'analyse de données sur l'expression des gènes (réponses continues) et aux données sur les modifications des histones (réponses discrètes de comptage).

Il s'agit d'un travail conjoint avec Fernando Quintana, Département de statistique de l'Université pontificale catholique du Chili.

. 15h20 à 15h50 : Pause-café

. 15h50 à 16h40 : Félix Camirand Lemyre, Professeur, Université de Sherbrooke

Estimation non paramétrique de l'apport habituel d'aliments épisodiquement consommés.

Dans des enquêtes nationales, il est souvent d'intérêt d'estimer la distribution de l'apport habituel de nutriments ou d'aliments. Dans ce contexte, les apports alimentaires sont souvent rapportés au moyen d'outils d'auto-évaluation qui mesurent la quantité d'aliments et de nutriments consommés dans une journée. Puisque cette mesure constitue un reflet imprécis de l'apport habituel, il est désormais reconnu que de telles observations peuvent être considérées comme des versions de l'apport habituel contaminées par des erreurs de mesure. Dans le cas d'un apport quotidien, il est possible d'utiliser les méthodes classiques d'erreurs de mesure pour estimer la distribution de l'apport moyen à partir de ce type de données. Or, ces méthodes ne peuvent plus être utilisées lorsque l'aliment est consommé épisodiquement puisque les données récoltées présenteront typiquement un excès de zéros. Nous abordons ici ce problème en adoptant une approche non paramétrique.

. 16h40 : Fin du Colloque