

7^e Rencontres Scientifiques Sherbrooke-Montpellier

Colloque de statistique et de biostatistique

Horaire et Programme

Toutes les présentations auront lieu au D4-2019

Mercredi le 19 juin 2019

10 h 30 : Ouverture du Colloque

SESSION 1 (Taoufik Bouezmarni)

10 h 30 à 11 h 15 : Jocelyn Chauvet, Doctorant, IMAG, Université de Montpellier. ***Régularisation des modèles linéaires généralisés mixtes par construction de composantes supervisées.***

11 h 15 à 12 h : Hanene Ben Salah, Institut supérieur de gestion de Tunis, Tunisie. ***Gestion des actifs Financiers : de l'approche Classique à la modélisation non paramétrique en estimation du downside Risk pour la Constitution d'un portefeuille efficient.***

12 h à 13 h 30 : Diner

SESSION 2 (Sévérien Nkurunziza)

13 h30 à 14 h 30 : Fátima Palacios Rodriguez, Postdoctorante, IMAG et INRIA, Montpellier. ***Rainfall analysis by Pareto processes.***

14 h 15 à 15 h : Aziz L'Moudden, Postdoctorant, Université de Sherbrooke et UQÀM. ***Une approche décisionnelle pour l'estimation de densités prédictives et l'analyse bayésienne.***

15 h à 15 h 30 : Pause santé (D3-1027-5)

SESSION 3 (Bernard Colin)

15 h 30 à 16 h 15 : Sévérien Nkurunziza, Professeur, Université Windsor (Ontario). ***Inférence dans les processus de retour à la moyenne généralisés avec plusieurs points de rupture.***

16 h 15 à 17 h : Tiffany Chefany, Doctorante, IMAG, Université de Montpellier. ***Politique de maintenance optimisée pour les systèmes multi-composants.***

17 h : Réception au Centre Culturel de l'Université de Sherbrooke

Jeudi 20 juin 2019 :

Rencontre estivale du Laboratoire de statistique et Deuxième journée du Colloque

SESSION IV. (Ali Gannoun)

9 h à 9 h 50 : Benjamin Charlier, Professeur, IMAG, Université de Montpellier. *Kernel Operations on the GPU, with autodiff, without memory overflows.*

9 h 50 à 10 h 20 : Pause-santé (D3-1027-5)

SESSION V. (Éric Marchand)

10 h 20 à 11 h 10: Mireille Schnitzer, Professeure, Université de Montréal. *Methodological considerations for the analysis of relative treatment effects in multi-drug-resistant tuberculosis from fused observational studies.*

11 h 10 à 12 h : Khader Khadraoui, Professeur, Université Laval. *Un seuillage adaptatif optimal pour la connectivité et la détection de communautés dans un réseau social aléatoire géométrique.*

12 h à 13 h 40 : Lunch

SESSION VI. (Félix Camirand Lemyre)

13 h 40 à 14 h 30 : 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.*

14 h 30 à 15 h 20 : Alejandro Murua, Professeur, Université de Montréal.

15 h 20 à 15 h 50 : Pause-santé (D3-1027-5)

SESSION VII (Benjamin Charlier)

15 h 50 à 16 h 40 : Félix Camirand Lemyre, Professeur, Université de Sherbrooke. *Estimation non paramétrique de l'apport habituel d'aliments épisodiquement consommés.*

16 h 40 : Fin du Colloque

Résumés

Jocelyn Chauvet : *Régularisation des modèles linéaires généralisés mixtes par construction de composantes supervisées.*

Résumé :

Une forte redondance des variables explicatives cause de gros problèmes d'identifiabilité et d'instabilité des coefficients dans les modèles de régression. Même lorsque l'estimation est possible, l'interprétation des résultats est donc extrêmement délicate. Il est alors indispensable de combiner à la vraisemblance de ces modèles un critère supplémentaire qui régularise l'estimateur. Dans le sillage de la régression PLS, la stratégie de régularisation que nous considérons est fondée sur l'extraction de *composantes supervisées*. Contraintes à l'orthogonalité entre elles, ces composantes doivent non seulement capturer l'information structurelle des variables explicatives, mais aussi prédire autant que possible les variables réponses, qui peuvent être de types divers (continues ou discrètes, quantitatives, ordinales ou nominales). La *régression sur composantes supervisées* a été développée pour les GLMs multivariés, mais n'a jusqu'alors concerné que des modèles à observations indépendantes. Or, dans de nombreuses situations, les observations sont groupées. Nous proposons une extension de la méthode aux GLMMs multivariés, pour lesquels les corrélations intra-groupes sont modélisées au moyen d'effets aléatoires. À chaque étape de l'algorithme de Schall permettant l'estimation du GLMM, nous procédons à la régularisation du modèle par l'extraction de composantes maximisant un compromis entre qualité d'ajustement et pertinence structurelle. Comparé à la régularisation par pénalisation de type ridge ou LASSO, nous montrons sur données simulées que notre méthode non seulement permet de révéler les dimensions explicatives les plus importantes pour l'ensemble des réponses, mais fournit souvent de meilleures estimations. La méthode est aussi évaluée sur des données réelles d'écologie forestière.

Hanene Ben Salah : *Gestion des Actifs Financiers: de l'Approche Classique à la Modélisation non Paramétrique en Estimation du DownSide Risk pour la Constitution d'un Portefeuille Efficient.*

Résumé :

Dans cette présentation, nous introduisons des estimateurs non paramétriques (basés sur la médiane ou l'espérance conditionnelles) pour remplacer les rendements des actions constituant un portefeuille et construire par ce biais une frontière efficiente lisse comparable à celle obtenue par des méthodes classiques développées par Yakowitz, le but final étant de construire un portefeuille optimal suivant certains critères.

Fatima Palacios Rodriguez : *Rainfall analysis by Pareto processes.*

Résumé :

Authorities aim to better understand and manage the risks of destructive natural disasters such as heavy rainfalls. For this purpose, impact models can be fed with simulations of extreme rainfall scenarios to study their sensitivity to temporal and spatial variability. The development of stochastic simulation methods incorporating realistic spatio-temporal extreme fields becomes crucial. In this work, we propose a semiparametric method that uses a moderate number of observed extreme space-time episodes to generate an unlimited number of space-time extreme scenarios of any magnitude. Our simulation framework draws sound theoretical justification from extreme value theory where space-time generalized Pareto processes arise as limits for event magnitudes exceeding a high threshold. Specifically, our implementation exploits asymptotic stability properties by decomposing extreme event episodes into a scalar magnitude variable (which is then resampled) and an empirical profile process representing space-time variation during an extreme episode. We illustrate this novel methodology on a high-dimensional gridded dataset of hourly precipitation data in Mediterranean France for events simulated according to yet unobserved magnitudes.

Aziz L'Moudden : *Une approche décisionnelle pour l'estimation de densités prédictives et l'analyse bayésienne.*

Résumé :

Les problèmes de prédiction sont au cœur de la statistique et interviennent dans plusieurs situations. Des défis existent quant au choix d'une densité prédictive efficace pour des données multidimensionnelles en particulier avec des procédures bayésiennes.

Lors de cet exposé, nous abordons l'estimation de densités prédictives et des mesures d'efficacité basées sur le risque fréquentiste. Notamment, pour les coûts Kullback-Leibler, de type alpha-divergence, nous démontrons que l'estimateur de Bayes associé à une mesure uniforme sur l'espace paramétrique contraint domine le meilleur estimateur invariant de la densité prédictive et nous obtenons des estimateurs minimax, nous présentons plusieurs résultats de dominance exploitant des techniques d'expansion d'échelle, des liens de dualité avec l'estimation et la prédiction ponctuelle, et l'estimation de Stein. Les modèles étudiés incluent la loi normale multivariée de structures de covariance connues, Gamma avec restriction ou information additionnelle sur les paramètres et des mélanges de lois.

Ce sont des travaux conjoints avec Éric Marchand de l'université de Sherbrooke.

Sévérien Nkurunziza : *Inférence dans les processus de retour à la moyenne généralisés avec plusieurs points de rupture.*

Résumé :

Dans cet exposé, nous présentons une méthode d'inférence dans les processus d'Ornstein-Uhlenbeck généralisés avec plusieurs points de rupture inconnus lorsque le paramètre dérive pourrait satisfaire une contrainte. L'un des traits saillant de la procédure établie réside dans le fait que le nombre de points de rupture ainsi que les emplacements des points de rupture sont inconnus. Nous affaiblissons les conditions d'existence des estimateurs sans (ES) et avec restriction (ER) et étudions leurs propriétés asymptotiques. De plus, nous établissons un test statistique qui permet de tester la véracité de la restriction hypothétique. Finalement, nous proposons une classe d'estimateurs à rétrécissement (SEs), qui comprend comme cas particuliers l'ES, ER ainsi que les SEs classiques. L'autre originalité de nos résultats asymptotiques réside dans le fait que les dimensions des estimateurs en question sont aléatoires. Pour ce faire, les méthodes asymptotiques classiques ne sont pas applicables

Tiffany Cherchi : *Politique de maintenance optimisée pour les systèmes multi-composants.*

Résumé :

Nous présentons un problème d'optimisation pour la maintenance d'un système multi-composants sujet à des détériorations ou défaillances aléatoires de ses composants, entraînant l'évolution de l'état général du système. Le système peut être requis pour effectuer des missions de fréquences et durées déterministes. Notre objectif à long terme est de mettre en place une politique de maintenance optimisée afin de garantir le bon déroulement des missions tout en minimisant les coûts de maintenance. L'idée principale de ce travail est de proposer un modèle mathématique pour l'évolution du système en utilisant le formalisme d'un Processus Markovien Décisionnel (MDP). Par simulations de Monte Carlo, nous comparons les performances de plusieurs politiques de référence.

Benjamin Charlier : *Kernel Operations on the GPU, with autodiff, without memory overflows.*

Résumé :

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.

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

Résumé :

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.

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

Résumé :

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.

Alejandro Murua : *Un modèle bayésien semi-paramétrique pour le regroupement des bi-grappes.*

Résumé :

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 Lazzeroni 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.

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

Résumé :

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.

Félix Camirand Lemyre : *Estimation non paramétrique de l'apport habituel d'aliments épisodiquement consommés.*

Résumé :

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.