

Réunion de rentrée
Campus de Beulieu, bat 27, salle
16 sept. 2016

Programme

9 :40	Accueil	
9 :45	Tour de table	
9 :50	Présentation des nouveaux doctorants	
	Florent Hébert/David Causeur	<i>Estimation et modélisation de la dépendance dans des tests d'association à grande échelle en génomique</i>
	Marie Morvan/Madison Giacofci	<i>Modèles de régression pour données fonctionnelles. Application à la modélisation de données de spectrométrie dans le proche infra rouge.</i>
	Arnaud Poinas/Bernard Delyon	<i>Statistique des processus déterminentaux.</i>
	Exposés	
10 :05	Guillaume Chauvet	<i>Théorème central-limite pour la méthode du pivot – application à l'échantillonnage spatial.</i>
10 :30	Pause thé et café	
11 :00	Pathé Ndao	<i>Estimation de quantiles extrêmes conditionnels en design aléatoire en présence de censure.</i>
11 :30	Audrey Poterie	<i>Arbres de décision pour les variables groupées.</i>
12 :00	Déjeuner	
13 :30	Gaspar Massiot	<i>Régression minimax pour une covariable poissonnienne.</i>
13 :55	Jean-François Dupuy	<i>Modèles de comptage avec zéro-inflation. Application en économie de la santé.</i>
14 :30	Pause thé et café	
	Réunion d'équipe	
15 :00		<i>Présentation du Master</i>
15 :30		<i>Point sur les groupes de travail, le séminaire</i>
15 :45		<i>Questions et informations diverses</i>

Résumés

Guillaume Chauvet, *Théorème central-limite pour la méthode du pivot – application à l'échantillonnage spatial*

Beaucoup de procédures d'échantillonnage à probabilités inégales existent dans la littérature. La méthode du pivot est particulièrement simple, et permet de sélectionner des échantillons bien répartis dans l'espace ce qui est particulièrement intéressant en cas de corrélation spatiale positive. Dans ce travail, nous prouvons un théorème central-limite de type martingale pour la méthode du pivot, et nous proposons également un estimateur de variance conservatif. Il s'agit d'un travail en collaboration avec Ronan Le Gleut (Insee).

Jean-François Dupuy, *Modèles de comptage avec zéro-inflation. Application en économie de la santé*

Zero-inflated regression models for count data are often used in health economics to analyse demand for medical care. Indeed, excess of zeros often affects health-care utilization data. We present some recent advances on the topic (asymptotic results in existing models and a new model), illustrated with numerical results and an application to analysis of health-care utilization.

Pathé Ndao, *Estimation de quantiles extrêmes conditionnels en design aléatoire en présence de censure*

Estimation of the extreme-value index of a heavy-tailed distribution is addressed when some random covariate information is available and the data are randomly right-censored. A weighted kernel version of Hill's estimator of the extreme-value index is proposed and its asymptotic normality is established. Based on this, a Weissman-type estimator of conditional extreme quantiles is constructed. A simulation study is conducted to assess the finite-sample behavior of the proposed estimators.

Audrey Poterie, *Classification trees for groups of variables*

In many situations, data sets consist of variables with an inherent group structure. In such cases, elaboration of a prediction rule according to the groups instead of the individual variables may be of interest for both interpretation and prediction accuracy. This problem has already been studied in regression (Group lasso [3], logistic group lasso [2],...). Yet, as far as we know, this problem has not been studied for decision trees and random forests. Recently, Gregorutti *et al.* adapted the permutation importance measure used in the random forests algorithm for groups of variables in order to perform a group variable selection [3]. Following this article, we suggest an algorithm elaborating decision trees based on group variables instead of individual variables. Our work is illustrated by a simulation study and an application on a dataset of patients with prostate cancer treated by radiotherapy.

[1] B. Gregorutti, B. Michel, and P. Saint-Pierre. Grouped variable importance with random forests and application to multiple functional data analysis. *Computational Statistics & Data Analysis*, 90 :15–35, 2015.

[2] L. Meier, S. L. Meier, S. Van De Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 70(1) :53–71, 2008.

[3] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(1) :49–67, 2006.