



14 MARS

🕒 09h00 à 11h00

📍 Campus Saint-Germain-des-Prés, U.
Paris Cité

THÈSES ET HDR

Myrto LIMNIOS : Processus de Rang et Applications Statistiques en Grande Dimension

Direction : Nicolas Vayatis et Ioannis Bargiotas

Soutenance le 14/03/2022

📅 AJOUTER AU
CALENDRIER

Titre

Processus de Rang et Applications Statistiques en Grande Dimension

Résumé

Ce projet de recherche propose de développer des outils mathématiques et algorithmiques pour étudier et comparer deux jeux de données complexes en grande dimension: vecteurs, signaux multivariés, trajectoires, signaux sur graphes. Il répond à des enjeux fondamentaux liés à la quantification dans les sciences expérimentales, notamment les sciences de la vie et par-là même les neurosciences et ses applications cliniques.

Pour se faire, nous proposons une généralisation des statistiques linéaires de rang à l'aide d'outils développés en apprentissage automatique. En effet, une étude approfondie de ces statistiques non-paramétriques à deux échantillons est menée sous l'angle de la théorie de l'apprentissage statistique. Plus précisément, des techniques d'ordonnement biparti permettent de pallier l'absence de relation d'ordre dans les espaces de grande dimension grâce à l'apprentissage d'une fonction de score. Définie sur l'espace ambiant et à valeur réelle, cette dernière a pour but d'induire un ordre sur les observations multivariées en maximisant la statistique de rang généralisée.

Nous proposons une première application dans le cadre des tests d'hypothèses statistiques, en associant décision (acceptation/rejet) de l'hypothèse nulle à l'apprentissage d'un modèle décrivant les données. Nous étudions, plus précisément, les tests d'homogénéité à deux échantillons. Ensuite, deux applications en analyse de données sont introduites et développées en utilisant les statistiques de rang comme critère scalaire de performance. Nous les appliquons aux problèmes d'ordonnement biparti et d'apprentissage des données extrêmes, ou anomalies, et précisons leurs relations à l'état de l'art. Enfin, dans la volonté de proposer des outils adaptés aux données issues des sciences expérimentales et dans le cadre de l'étude des données biomédicales, nous introduisons une méthode interprétable de comparaison statistique de deux populations cliniques, ainsi que d'un modèle stochastique génératif de données longitudinales particulières.

Direction

Nicolas Vayatis et Ioannis Bargiotas

Jury

- Mme Alexandra Carpentier, Professeure, Institut de mathématiques, Université de Potsdam, Allemagne [Rapporteur & Examinatrice]
- M. Johan Segers, Professeur, Institut de statistique, biostatistique et sciences actuariales, LIDAM, UCLouvain, Belgique [Rapporteur & Examineur]
- M. Stephan Cléménçon, Professeur, Telecom Paris, LTCI, Institut Polytechnique de Paris, France [Examineur]
- M. Pascal Massart, Professeur, Université Paris-Saclay, CNRS, Inria, Laboratoire de mathématiques d'Orsay, France [Examineur]
- Mme Sara van de Geer, Professeure, Séminaire de statistiques, Département de mathématiques, ETH Zurich, Suisse [Examinatrice]
- M. Nicolas Vayatis, Professeur, Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, France [Directeur]
- M. Ioannis Bargiotas, Chercheur, Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, France [Invité]

