

AVIS DE SOUTENANCE DE THÈSE

DOCTORAT (Arrêté du 26 août 2022 modifiant l'arrêté du 25 mai 2016)

Monsieur Tomasz SKALSKI

candidat au diplôme de Doctorat de l'Université d'Angers, est autorisé à soutenir publiquement sa thèse

le 13/10/2023 à 12h15

POLITECHNIKA WROCLAWSKA (WUST)
WROCLAW
POLOGNE

sur le sujet suivant :

Geometric and Combinatorial Aspects of Statistical Models

Directeur de thèse : **Monsieur Piotr GRACZYK**

Co-Directeur de thèse : **Monsieur Maciej WILCZYNSKI**

Composition du jury :

Monsieur Loïc CHAUMONT, Professeur des Universités Université d'Angers, Examineur
Monsieur Piotr GRACZYK, Professeur des Universités Université d'Angers, Directeur de thèse
Madame Alicja JOKIEL-ROKITA, Professeure Politechnika Wroclawska, Pologne, Examineur
Monsieur Fabien PANLOUP, Professeur des Universités Université d'Angers, Examineur
Monsieur Mateusz KWASNICKI, Professeur Politechnika Wroclawska, Pologne, Examineur
Monsieur Jan MIELNICZUK, Professeur Politechnika Warszawska, Pologne, Rapporteur
Monsieur Joseph SALMON, Professeur des Universités Université de Montpellier, Rapporteur
Monsieur Grzegorz SERAFIN, Dr Hab. Politechnika Wroclawska, Pologne, Examineur
Monsieur Gilles STUPFLER, Professeur des Universités Université d'Angers, Examineur
Monsieur Patrick TARDIVEL, Maître de Conférences Université de Bourgogne, Co-encadrant
Madame Sara VAN DE GEER, Professeure Ecole Polytechnique Fédérale de Zurich, Suisse, Rapporteur
Monsieur Piotr WIECEK, Professeur Politechnika Wroclawska, Pologne, Examineur
Monsieur Maciej WILCZYNSKI, Professeur Politechnika Wroclawska, Pologne, Co-directeur de thèse

Résumé de la thèse

Cette thèse traite des applications de la combinatoire et de la géométrie discrète aux statistiques modernes. La première application porte sur l'une des approches fréquemment utilisées pour faire face au volume croissant des données, à savoir l'utilisation de méthodes de régression linéaire pénalisée. Dans le but de retrouver la structure que possède le vecteur des coefficients de régression, nous commençons notre discussion par l'estimateur SLOPE (Sorted l1 Penalized Estimator), qui a été proposé il y a presque dix ans. En particulier, nous examinons la notion de schéma SLOPE, qui conserve l'information sur le support, le signe et le classement des coefficients de régression. En particulier, il préserve les groupes de coefficients ayant la même valeur absolue. Dans le chapitre 3, nous fournissons les conditions, sous lesquelles SLOPE récupère l'ensemble des variables pertinentes et des groupes lorsque la matrice de planification est orthogonale. Nous déduisons également de nouveaux résultats sur la forte consistance de l'estimateur SLOPE et de son schéma. Le chapitre 4 étend la discussion sur l'estimateur SLOPE à une classe générale de matrices de planification fixes. Nous fournissons la condition d'irreprésentabilité du SLOPE, qui est nécessaire et suffisante pour le recouvrement du schéma dans le cadre non-bruité et nous illustrons cette propriété géométriquement. Dans un deuxième temps, nous considérons le cas asymptotique lorsque le nombre de variables explicatives tend vers l'infini et que l'erreur est incrémentale. Dans le chapitre 5, nous étudions la classe plus large des estimateurs pénalisés, appelés les jauges polyédriques. Elle permet d'utiliser les notions issues de la géométrie des polyèdres pour généraliser la notion du schéma et les résultats sur sa récupération. Le chapitre 6 s'articule autour de l'existence de l'estimateur du maximum de vraisemblance (MLE) pour les familles exponentielles discrètes. Nous donnons sa nouvelle caractérisation basée sur la notion d'ensemble d'unicité. Par la suite, nous inspectons la taille des échantillons indépendants identiquement distribués qui est nécessaire pour assurer son existence avec une grande probabilité. Pour cela, nous utilisons les notions issues de l'analyse des hypercubes discrets et appliquons nos résultats dans l'environnement des graphes aléatoires. Le dernier des chapitres relie la théorie des schémas graphiques en statistique avec la notion de matrices laplaciennes de graphes et de processus de Wiener discrétisés. La thèse est basée sur trois articles déjà publiés et deux prépublications, qui sont disponibles en ligne.

À AFFICHER DANS L'UFR 15 JOURS AVANT LA SOUTENANCE