

AVIS DE SOUTENANCE DE THÈSE

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

Monsieur Maxime EGÉA

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

le 29/03/2023 à 10h00

Faculté des Sciences

AMPHI L001

2, boulevard Lavoisier

49045 ANGERS Cedex 01

sur le sujet suivant :

Méthodes multilevel pour l'apprentissage bayésien et applications aux données biologiques

Directeur de thèse : **Monsieur Fabien PANLOUP**

Composition du jury :

Monsieur Arnak DALALYAN, Professeur des Universités ENSAE Palaiseau, Examineur

Monsieur Andreas EBERLE, Professeur Université de Bonn, Allemagne, Rapporteur

Madame Gersende FORT, Directrice de Recherche CNRS Université Paul Sabatier, Toulouse, Examineur

Monsieur Ahmed KEBAIER, Professeur des Universités Université d'Evry, Rapporteur

Monsieur Gilles PAGÈS, Professeur des Universités Sorbonne Université, Examineur

Monsieur Fabien PANLOUP, Professeur des Universités Université d'Angers, Directeur de thèse

Résumé de la thèse

Cette thèse est issue d'un partenariat entre le laboratoire de Mathématiques d'Angers (LAREMA) et le SIRIC-ILIAD (Site de Recherche Intégrée sur le Cancer) Nantes-Angers. L'objectif est le développement de méthodes numériques efficaces pour l'apprentissage bayésien des données inhérentes aux cancers. Outre la difficulté statistique liée au faible nombre d'individus comparé au nombre de données acquises par patient, la grande dimension impact fortement l'efficacité des méthodes numériques et nécessite le développement de nouvelles méthodes capables d'appréhender cette même dimension. Après une introduction visant à présenter les outils et résultats existant, le premier travail de cette thèse introduit un nouvel algorithme multilevel. D'abord décrite dans un cadre général, la complexité de cet algorithme est calculée plus précisément pour des diffusions de Langevin satisfaisant des hypothèses d'uni forme convexité. En plus de décrire explicitement les dépendances en la dimension, ces résultats améliorent ceux existant dans la littérature. Dans un deuxième temps, on cherche à affaiblir l'hypothèse d'uniforme convexité afin de mieux répondre à certains objectifs statistiques. Dans ce cadre difficile, deux techniques sont étudiées. Dans la première, l'idée est d'ajouter une composante fortement convexe au potentiel faiblement convexe afin de pouvoir utiliser les résultats de la première partie. Dans la seconde partie, on se place dans un cadre intermédiaire entre l'uniforme et la faible convexité. Après avoir montré des résultats concernant les moments exponentiels ou encore la distance en temps long à la diffusion, on montre que l'algorithme multilevel s'adapte à ce cadre et permet d'approcher l'estimateur avec une complexité explicite en les paramètres. Le dernier travail consiste à illustrer l'application des méthodes précédemment évoquées aux données réelles. Ainsi, on étudie un jeu de données contenant des données génomiques de patient atteint de cancer du sein. Après une réduction de la dimension, on calcule la moyenne a posteriori associée à une régression logistique visant à prévoir la réponse aux traitements.