Mini-symposium SG
Quelques connexions entre Statistique et Géométrie

Mini-symposium porté par l'ANR TopData

Résumé

Dans le contexte du développement de l'Analyse Topologique des Données (TDA), domaine actuellement en plein essor, ce mini-symposium est consacré aux liens entre statistique et géométrie. L'objectif est de fournir un aperçu de différents travaux récents en statistique comportant tous une forte composante géométrique. En effet, dans les problèmes modernes d'analyse des données et d'apprentissage statistique, on a affaire de plus en plus fréquemment à des observations complexes et/ou de grande dimension. Ainsi, il est nécessaire de disposer d'outils adaptés permettant de comprendre et d'exploiter les structures géométriques sous-jacentes. Souvent, ces problèmes conduisent à adopter une démarche combinant statistique, topologie, géométrie et algorithmique (cf aussi le mini-symposium *Géométrie numérique et algorithmique* organisé par Boris Thibert).

Organisateur(s)

1. **Aurélie Fischer**, Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot.

Liste des orateurs

1. **Claire Brécheteau**, Université Paris-Sud et INRIA Saclay DataShape
   Titre : Un test d'isomorphisme entre espaces métriques mesurés utilisant la distance à la mesure.
2. **Sylvain Delattre**, Université Paris Diderot
   Titre : Sur les courbes principales de longueur bornée.
3. **Ilaria Giulini**, Université Paris Diderot et INRIA Saclay DataShape
   Titre : Estimation automatique de l'opérateur de Laplace-Beltrami.
4. **Clément Levrard**, Université Paris Diderot
   Titre : Inférence d'objets différentiels par interpolation polynomiale : vitesses minimax pour l'estimation de support, plans tangents et courbure.

Aurélie Fischer, Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot, Case courrier 7012, Place Aurélie Nemours, 75205 Paris Cedex 13, `aurelie.fischer@univ-paris-diderot.fr`
Claire Brécheteau, Centre Inria Saclay - Île-de-France Bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves, Campus de l'École Polytechnique, 91120 Palaiseau, `claire.brecheteau@inria.fr`
Sylvain Delattre, Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot, Case courrier 7012, Place Aurélie Nemours, 75205 Paris Cedex 13, `sylvain.delattre@univ-paris-diderot.fr`
Ilaria Giulini, Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot, Case courrier 7012, Place Aurélie Nemours, 75205 Paris Cedex 13 & Centre Inria Saclay - Île-de-France Bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves, Campus de l'École Polytechnique, 91120 Palaiseau, `ilaria.giulini@inria.fr`
Clément Levrard, Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot, Case courrier 7012, Place Aurélie Nemours, 75205 Paris Cedex 13, `clement.levrard@univ-paris-diderot.fr`

Introduction

De nombreux domaines d'applications fournissent de nos jours des observations de nature complexe et/ou de grande dimension. Pour extraire de l'information pertinente à partir de telles données, il est important de comprendre leur structure géométrique sous-jacente. Dans ce cadre, les exposés de la session, qui se concentreront avant tout sur les fondements théoriques des méthodes d'apprentissage envisagées, proposent de présenter quelques connexions entre statistique et géométrie. Les sections qui suivent fournissent une description de chacun des exposés.

1 Un test d'isomorphisme entre espaces métriques mesurés utilisant la distance à la mesure

Dans cet exposé sera introduite la notion de signature DTM, mesure sur $\mathbb{R}_+$ pouvant être associée à tout espace métrique mesuré. Cette signature est basée sur la distance à la mesure introduite dans [3]. Elle conduit à une pseudo-distance entre espaces métriques mesurés, qui est majorée par la distance de Gromov-Wasserstein. Sous certaines conditions géométriques, des bornes inférieures peuvent être obtenues pour cette pseudo-distance. De plus, étant donné deux n-échantillons, il est possible, en utilisant la signature DTM, de construire un test statistique asymptotique d'égalité des espaces métriques sous-jacents, modulo isométries préservant la mesure.

2 Sur les courbes principales de longueur bornée

Les courbes principales sont des courbes paramétrées passant au milieu d'une loi de probabilité dans $\mathbb{R}^d$. Outre la définition originelle basée sur la notion d'auto-consistance, plusieurs points de vue ont été considérés, dont un problème de minimisation de type moindres carrés avec contrainte. Dans cet exposé, nous étudions les propriétés théoriques de courbes principales de longueur au plus L et montrons notamment qu'elles sont toujours de courbure finie.
A partir de la condition d'ordre 1, exprimant qu'une courbe est un point critique pour le critère, nous obtenons une équation faisant intervenir la courbe, sa courbure, ainsi qu'une variable aléatoire jouant le rôle du paramètre de la courbe paramétrée. Cette équation permet de proposer une nouvelle démonstration de l'injectivité d'une courbe principale contrainte en dimension 2.

3 Estimation automatique de l'opérateur de Laplace-Beltrami

L'estimation de l'opérateur de Laplace-Beltrami sur une variété joue un rôle important dans les méthodes spectrales d'analyse des données, comme par exemple le clustering spectral. Cette estimation se fait généralement à l'aide d'un estimateur à noyau qui dépend d'un paramètre d'échelle. La bonne calibration de ce paramètre est essentielle en pratique mais établir une méthode rigoureuse pour ce choix reste un problème ouvert à ce jour. Dans cet exposé, nous proposons un premier pas pour répondre à cette question, en nous appuyant sur des techniques de Lepski et des résultats de [1].

4 Inférence d'objets différentiels par interpolation polynomiale : vitesses minimax pour l'estimation de support, plans tangents et courbure.

L'approximation de sous-variétés riemanniennes par des polynômes locaux est une technique courante en reconstruction de surfaces ou de courbes. Plusieurs résultats déterministes sur la qualité d'approximation de quantités telles que la courbure et les plans tangents ont été détaillés dans [2]. Dans cet exposé, nous donnons des extensions de ces résultats au cadre d'un échantillonnage probabiliste, en nous attachant à mettre en lumière la dépendance des vitesses de convergence en les dimensions (intrinsèque et ambiante) et en la taille d'échantillon, démontrant que la dimension ambiante ne joue quasiment aucun rôle. Ces vitesses sont étayées par l'obtention de résultats minimax sur la classe d'objets (sous-variétés) considérée.

Références

[1] C. Lacour, P. Massart, V. Rivoirard, *Estimator selection : a new method with applications to kernel density estimation*, 2016, prépublication.

[2] F. Cazals, M. Pouget, *Estimating Differential Quantities Using Polynomial Fitting of Osculating Jets*, L. Kobbelt, P. Schröder, H. Hoppe. Eurographics Symposium on Geometry Processing, 2003, 177–187.

[3] F. Chazal, D. Cohen-Steiner, Q. Mérigot, *Geometric Inference for Probability Measure*, Foundations of Computational Mathematics, 2011, 11 : 733–51.