1

Méthodes adaptatives

Session organisée par Patricia Reynaud-Bouret

Le but de la session est de présenter quelques méthodes d'estimation adaptative, en particulier de montrer l'importance des inégalités oracles dans divers contexte : apprentissage, prédiction de suite individuelles, régression, modèles spatiaux (1).

Références :

[1] *Lectures from the 33nd Probability Summer School held in Saint-Flour, July 7–24, 2003, Edited by Jean Picard, to appear,* (2005) Pascal Massart, Springer-Verlag.

Adresse de l'organisateur :

Patricia Reynaud-Bouret
Département de mathématiques et applications
ENS Paris
45 rue d'Ulm
75230 Paris Cedex 05, France
E-mail : `Patricia.Reynaud@ens.fr`
<`http://www.dma.ens.fr/~reynaud/`>

Le marteau d'Occam

par **Gilles Blanchard**

quel est le rapport entre l'estimation ou la sélection de modèles randomisées et le contrôle du 'False Discovery Rate' en test d'hypothèses multiples ? Une réponse peut-être inattendue est : le théorème de Fubini. Nous montrons une inégalité abstraite (surnommée affectueusement "marteau d'Occam") s'appliquant à tout algorithme statistique retournant un sous-ensemble d'un ensemble d'objets H fixés au départ (estimateurs, hypothèses nulles,...) ou plus généralement une distribution sur H. Celle-ci peut se voir comme une généralisation de la borne d'union et utilise essentiellement le théorème de Fubini. Nous expliquerons en quoi elle peut s'appliquer aux deux exemples cités ci-dessus pour donner des conclusions intéressantes.

Adresse :

Gilles Blanchard
Fraunhofer FIRST (IDA)
Kékuléstrasse 7
12489 Berlin, Allemagne
E-mail : blanchar@first.fhg.de
<http://www.math.u-psud.fr/~blanchard/>

Prédiction de suites individuelles : estimation adaptative sans distribution sous-jacente

par **Gilles Stoltz**

La prédiction des suites individuelles considère les problèmes d'apprentissage séquentiel pour lesquels on ne peut où ne veut pas modéliser le problème de manière stochastique, et fournit dès lors des stratégies de prédiction très robustes. Elle englobe aussi bien des problèmes issus de la communauté du machine learning que de celle de la théorie des jeux répétés. Ces derniers sont traités avec des méthodes statistiques, incluant par exemple les techniques de concentration de la mesure, de la théorie de l'information, mais aussi celles de l'estimation adaptative. Les similitudes avec l'estimation adaptative concernent autant les methodes (nécessité de régler des paramètres des algorithmes en fonction de quantités inconnues sur les fonctions d'intérêt) que les résultats obtenus, formulés, en un certain sens, commes des inégalités oracles exactes sans distribution sous-jacente.

Adresse :

Gilles Stoltz
Département de mathématiques et applications
ENS Paris
45 rue d'Ulm
75230 Paris Cedex 05, France
E-mail : Gilles.Stoltz@ens.fr
<http://www.dma.ens.fr/~stoltz/>

Sélection de modèles linéaires par seuillage aléatoire

par **Marc Lavielle** et Carin Ludena

Une nouvelle méthode est proposée pour estimer le nombre de coeÿicients significatifs dans un problème de sélection de modèles. Cette méthode utilise un centrage aléatoire bien choisi des sommes cumulées partielles des observations ordonnées. En utilisant les propriétés des L-statistiques, nous montrons la consistance de l'estimateur proposé. La procédure est ensuite appliquée à un modèle de régression et est interprétée comme une procédure de seuillage aléatoire. Des exemples numériques illustrent l'intérêt pratique de la méthode. Si le temps le permet, nous montrerons comment étendre cette approche à des problèmes de sélection de modèles plus généraux (distributions paramétriques inconnues, détection de ruptures ...).

Adresses :

Marc Lavielle
Université René Descartes et INRIA Futurs, Projet SELECT
Université Paris-Sud
Bâtiment 425
91405 Orsay Cedex, France
E-mail : `Marc.Lavielle@math.u-psud.fr`
<`http://www.math.u-psud.fr/~lavielle/`>

Carin Ludena
IVIC,
Caracas, Venezuela
E-mail : `cludena@ivic.ve`

Sélection de modèles de fonctions polynomiales par morceaux en régression

par **Marie Sauvé**

Notre travail traite du problème de la sélection de modèles, par minimisation d'un critère des moindres carrés pénalisés, en régression. Notre but est de généraliser le résultat de Birgé et Massart en régression gaussienne au cas où les erreurs sont seulement supposées avoir des moments exponentiels au voisinage de 0. Nous mettons en évidence les difficultés d'une telle généralisation. Nous exhibons une forme de pénalité qui permet de sélectionner un modèle de fonctions polynomiales par morceaux dont l'estimateur des moindres carrés associé vérifie une inégalité de type oracle (2; 3).

Références :

[2] *Gaussian model selection* (2001) Lucien Birgé et Pascal Massart, JEMS.
[3] *Minimal penalties for gaussian model selection,* (2005) Lucien Birgé et Pascal Massart, à paraître dans PTRF.

Adresse :

Marie Sauvé
Laboratoire de Mathématiques
Université Paris-Sud
Bâtiment 425
91405 Orsay Cedex, France
E-mail : `Marie.Sauve@math.u-psud.fr`

Sélection de voisinage pour des champs de Markov gaussiens

par **Nicolas Verzelen**

Dans cet exposé, on s'intéressera à l'étude de données spatiales régulièrement réparties en utilisant des champs markoviens gaussiens. Si on connaît le voisinage du graphe sous-jacent, il existe ce nombreuses techniques d'estimation de la loi jointe. Cependant l'estimation du voisinage se révèle beaucoup plus problématique.

Après avoir rappelé les notions de bases sur les champs markoviens gaussiens, on s'attachera à présenter une procédure de sélection de modèle pour estimer en même temps le voisinage et les autres paramètres du champ. Cette méthode s'appuiera sur des techniques d'estimation par minimisation de contraste pénalisé. Enfin, nous appliquerons ces résultats à des données environnementales.

Adresse :

Nicolas Verzelen
Laboratoire de Mathématiques
Université Paris-Sud
Bâtiment 425
91405 Orsay Cedex, France
E-mail : `Nicolas.Verzelen@ens.fr`
<`http://www.eleves.ens.fr/home/nverzele/`>