

# Estimation d'un modèle de régression semi-paramétrique contaminé

Pierre VANDEKERKHOVE, Université Paris-Est Marne-la-Vallée

**Mots-clés** : mélange, régression, identifiabilité semi-paramétrique, processus empirique.

Nous nous intéressons dans cette communication à un modèle dit de *régression contaminé*. Ce modèle correspond à un mélange de régression à deux composantes (voir Martin Magniette *et al.* [2] pour une application de ce type de modèle en analyse des puces ADN), dans lequel un des modèles de régression est entièrement connu.

Notre modèle est dit semi-paramétrique au sens où la densité de probabilité du bruit impliqué dans le modèle de régression *inconnu* n'est pas supposée appartenir à une famille paramétrique. Lorsque les densités du bruit inconnu et du phénomène contaminant sont supposées symétriques autour de zéro, nous proposons un estimateur consistant des divers paramètres du modèle (euclidiens et fonctionnels). Nous montrons en particulier, sous des hypothèses faibles toutes satisfaites lorsque le modèle sous-jacent est gaussien, que la partie euclidienne, respectivement fonctionnelle, du modèle peut être estimée presque sûrement à la vitesse  $o(n^{-1/4+\alpha})$ , resp.  $o(n^{-1/6+\alpha})$  (au sens de la norme uniforme sur les fonctions de répartition), pour tout  $\alpha > 0$ . Comme dans Bordes et Vandekerkhove [1], si un TCL fonctionnel peut être obtenu pour nos paramètres (ce qui semble raisonnable), le résultat que nous venons d'énoncer en sera un des arguments clé. Nous étudierons enfin les performances numériques de notre estimateur sur divers exemples test. Nous pourrons à cette occasion observer l'existence d'artefacts d'estimation préalablement mis en évidence sur le plan théorique.

## Références

- [1] L. BORDES, P. VANDEKERKHOVE., *Semiparametric two-component mixture model when a component is known: an asymptotically normal estimator*, Mathematical Methods of Statistics, **19**, p. 22-41, 2010.
  
- [2] MARTIN-MAGNIETTE, M-L., MARY-HUARD, T., BERARD C. AND ROBIN, S., *ChIPmix: mixture model of regressions for two-color ChIP-chip analysis*, Bioinformatics, **24**, p. 181-186, 2008.