

**METHODES DE REGRESSION SEMIPARAMETRIQUE  
DE TYPES « SLICING » OU « POOLED SLICING » :  
MISE EN ŒUVRE SOUS LE LOGICIEL SAS®  
SOUS FORME DE MACRO-COMMANDES  
ET APPLICATION SUR DES JEUX DE DONNEES**

**Ali Gannoun<sup>1,2</sup>, Christiane Guinot<sup>3</sup> et Jérôme Saracco<sup>1</sup>**

<sup>1</sup> *Laboratoire de Probabilités et Statistique  
Université Montpellier II  
Place Eugène Bataillon, 34 095 Montpellier Cedex 5, France  
(Email : {gannoun,saracco}@stat.math.univ-montp2.fr)*

<sup>2</sup> *Statistical Genetics and Bioinformatics unit  
National Human Genome Center  
Howard University  
Washington D.C. 200059, U.S.A.  
(Email : agannoun@howard.edu)*

<sup>3</sup> *CE.R.I.E.S.  
20, rue Victor Noir, 92 521 Neuilly sur Seine Cedex, France  
(Email : christiane.guinot@ceries-lab.com)*

## 1. Introduction

De nombreuses méthodes de régression et de prédiction sont disponibles quand la variable à expliquer  $Y \in \mathbf{R}$  est réelle et lorsque la variable explicative  $X$  est un vecteur de  $\mathbf{R}^p$ . Ces méthodes reposent soit sur une approche paramétrique (comme la régression linéaire multiple ou polynomiale), soit sur des approches non-paramétriques (comme la régression à noyau ou les splines de lissage) ou semi-paramétriques (telles que celles que nous allons présenter ici). Toutes ces approches et méthodes sont caractérisées par des avantages et des défauts spécifiques.

Ainsi, par exemple, dans les modèles paramétriques, la structure de la dépendance entre  $Y$  et  $X$  est complètement fixée. Pour l'interprétation du modèle, ceci est un point très positif. Cependant, dans certains cas, le modèle paramétrique n'est pas en adéquation avec le vrai modèle sous-jacent ; la méthode estime alors la « meilleure » approximation dans l'espace de la classe choisie de fonctions paramétriques. Cette approximation peut être parfois très « éloignée » du vrai modèle, et les conclusions découlant de l'estimation peuvent alors être erronées.